A MACHINE-LEARNING APPROACH TO CHANGE DETECTION USING MULTI-SCALE IMAGERY[1]

Lisa M. Levien
Remote Sensing Specialist
USDA Forest Service
1920 20th Street
Sacramento, CA 95814
llevien/r5_rsl@fs.fed.us

Peter Roffers, Barbara Maurizi, and James Suero
Remote Sensing Analysts
Pacific Meridian Resources
1920 20th Street
Sacramento, CA 95814
roffers@cdf.ca.gov, bmaurizi@hoot.cdf.ca.gov, and james@cdf.ca.gov

Chris Fischer
Remote Sensing/GIS Specialist
California Department of Forestry and Fire Protection
1920 20th Street
Sacramento, CA 95814
chris_fischer@fire.ca.gov

Xueqiao Huang
Principal Information Management Specialist
ACS Government Solutions Group, Inc.
New Orleans, LA 70123
xueqiao.huang@mms.gov

The USDA Forest Service and the California Department of Forestry and Fire Protection are collaborating on a statewide change detection program to identify landcover change across all ownerships within five-year time periods. This program uses Landsat Thematic Mapper satellite imagery to derive landcover change and aid in assessing its cause. Landscape changes are initially detected using a multi-temporal Kauth-Thomas transform. Unsupervised classification of the transformed imagery creates a preliminary landscape-level change map portraying change classes with multiple levels of vegetation increase, decrease, and no change. Using a stratified random sampling scheme, this preliminary change map facilitates selection of field sites for collecting vegetation canopy cover measurements. Ground truth for the classifier is obtained by estimating canopy cover change over the 5-year timeframe using color-IR digital photos, digital orthophoto quads, and aerial photography. Canopy cover estimates from the second date of photography are calibrated using transect measurements of canopy cover from a sample of field sites. Attributes such as species and vegetative cover are also noted. A machine learning classifier approach is then employed. The classifier uses an inductive learning algorithm to generate production rules from training data, including the transformed change data and other ancillary data layers. The resultant knowledge base is then used by an expert classifier to produce classes of crown closure change. Approximately half of the field sites are reserved for accuracy assessment.

INTRODUCTION

Change detection is the process of identifying differences in landcover over time. As human and natural forces continue to alter the landscape, various public agencies are finding it increasingly important to develop monitoring methods to assess these changes. Changes in vegetation result in changes in wildlife habitat, fire conditions, aesthetic and historical values, ambient air quality, and other resource values, which in turn influence policy decisions. Currently, the USDA Forest Service is interested in assessing drought-and insect-caused mortality within

[1] Presented at the American Society of Photogrammetry and Remote Sensing 1999 Annual Conference, Portland, Oregon, May 20, 1999.

coniferous forests, while State and other local agencies are concerned with the loss of oak woodlands to conversion, fuelwood harvest and urban development.

Methods for monitoring vegetation change at a landscape scale range from fieldwork intensive plot inventories to utilization of remotely sensed data that include aerial photography and satellite imagery. The USDA Forest Service (FS), in collaboration with the California Department of Forestry and Fire Protection (CDF), is conducting a statewide change detection program using Landsat Thematic Mapper (TM) satellite imagery in an effort to improve monitoring data quality and minimize monitoring costs. The program goal is to identify vegetative change over a five year time frame statewide. The state has been divided up into five project areas, and each project area will take one year to complete.

BACKGROUND

Satellite imagery and image processing to monitor changes in landcover has been well documented over the past 20 years. Previous studies conducted cooperatively with FS and Boston University (BU) assessed the utility of using remote sensing for monitoring and mapping conifer mortality. In 1991 an initial study was undertaken to evaluate the potential for measuring and mapping conifer loss due to drought-related effects using TM imagery (Macomber and Woodcock, 1994). The approach, an extension of a vegetation mapping project, used a geometric canopy reflectance model to determine canopy closure and tree size (Woodcock, et al., 1994). The assumption was that changes in canopy closure could be mapped and related to changes due to mortality. Intensive field work was required to collect enough data on the ground to correlate it with the canopy model. This investigation was conducted on the Lake Tahoe Basin Management Unit in the Sierra Nevada.

In a later study from 1994 to 1995, FS and CDF entered into another cooperative agreement with BU to develop a more simplified methodology for implementing change detection over large areas to monitor conifer mortality. After investigating various methods, a multi-temporal Kauth-Thomas (MKT) transformation was chosen (Collins and Woodcock, 1996). The MKT differencing was shown to be very sensitive to change, and the regression technique used required much less field work.

The cooperative established between FS and CDF has opened up the possibility of exploring various applications of this methodology. CDF is interested in the hardwood rangelands throughout the state of California, and this method shows promise as a monitoring tool for evaluating declines in hardwood canopy cover due to fire, thinning, harvest, urban development, and other factors.

PROJECT AREAS

Project areas are initially determined by the boundaries of Landsat scenes and ecological subsections from the National Hierarchical Framework of Ecological Units. Two project areas including the southern Sierra and the northern Sierra have been completed with this initial division of the state. In FY98 a new coordinated schedule with slightly different project area boundaries was implemented. The areas within the coordinated schedule were determined by Landsat scene coverage and ecological subsection boundaries, and include approximately 5 million acres gross of national Forest System lands. Figure 1 depicts the current coordinated schedule project area map.

The coordinated schedule was developed to acquire aerial photography and TM imagery in the summer before the year of the change detection effort. This will be followed by vegetation mapping updates on National Forests and other state and private ownerships, forest inventory remeasurements of changed areas, and follow-up analysis. The coordinated schedule realizes cost savings by multiple programs and ensures a five year cycle of updating and monitoring throughout the state.

Project areas range in size from approximately 16 - 20 million acres in size, covering all ownerships. Our first test of the machine-learning methodology was performed on a pilot area (350,000 acres) within the southern California project area, which covers 17 million acres. Figure 2 shows the location of the southern California project area and pilot study area.

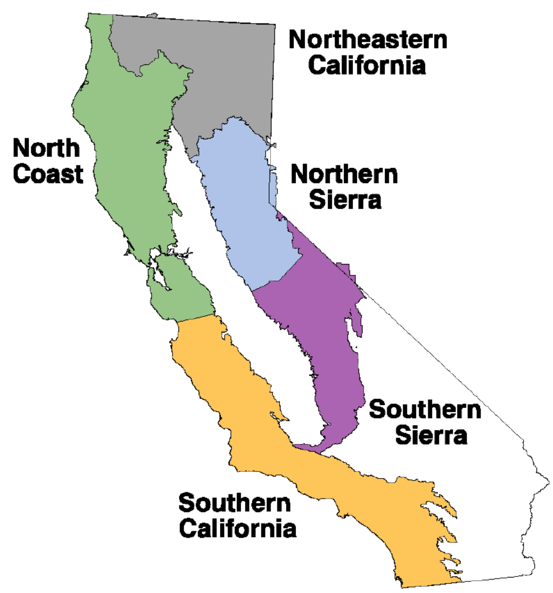

Figure 1. Change detection project areas in California.

Figure 2. Pilot study area within the Southern California project area.

METHODS

Phase I (Figure 3)

When using satellite imagery to detect change, imagery must be radiometrically corrected and co-registered. Image registration ensures that multidate images from the same path and row are registered to each other within one-half pixel by on-screen identification of common features, such as road intersections. If pixels do not correctly correspond, then changes due to misregistration will occur on the final change map. Geometric correction was performed for all image pairs in the project area. Correction of each image pair involved approximately 50 ground control points, producing an average root mean square error of 12.6 meters (8.9 m in the X direction and 8.9 m in the Y direction).

Radiometric correction ensures that any detected changes are not the result of differences in atmospheric conditions between the two dates of imagery. Atmospheric correction for this project involves converting original digital number (DN) values to reflectance. This is done by modeling the contribution of the atmospheric path radiance and absorption to the measured signal, and then removing those influences. While simpler methods of atmospheric correction have been shown suitable for change detection studies, this method was chosen in order to normalize multiple adjacent Landsat TM scenes and allow the machine-learning classifier to work effectively across scenes. DN values are converted to radiance and input to the software package 6S (Tanre, et al., 1990), which models the atmospheric parameters necessary for conversion of radiance to reflectance. These parameters are then applied to the imagery to perform the correction.

Imagery that has been normalized, registered, and subset into processing areas is ready for input into the change detection process. A concurrent process involves preparing and mosaicking ancillary data layers, including vegetation, fire history, plantation, and other harvest information. Ancillary data are used both as a masking tool and as a means for stratification to label the change classes and implement the sampling design for field data collection.

Change processing involves image segmentation and MKT transformation. Image segmentation creates regions (polygons based on spectral similarity) from TM bands 3 and 4, and a texture band generated from band 4 (Ryherd and Woodcock, 1990). Texture is a spatial component that enhances subtle edges in the scene over large areas. Generally, regions ranged from 15 to 50 acres. The MKT transform is a linear transformation that reduces several TM bands into brightness, greenness, and wetness components. Brightness identifies variation in reflectance, greenness is related to the amount of green vegetation present in the scene, and wetness correlates to canopy and soil moisture. The MKT transform is applied to the two dates of imagery. The difference is a change image representing the difference in brightness, greenness, and wetness values between the two dates.

The final step of the Phase I process is to identify change classes based on the change image. The change data are stratified by lifeform (e.g., conifer, hardwood, shrub, grass, non-forested/other) using the mosaicked vegetation data layer. An unsupervised classification is applied to each change image by lifeform and results in 50 change classes per lifeform. Image appearance, photo interpretation, vegetation and topographic maps, GIS coverages, and bispectral plots (e.g., greenness vs. wetness) assists in identifying levels of change. Each change class is labeled according to its level of change based on a gradient of change classes from large decreases in vegetation to large increases in vegetation (Table 1, left side, and Figure 4).

Phase II (Figure 5)

A machine-learning classifier is used to create the final change map based on classes of canopy cover change. This type of quantitative classification scheme is considered more useful to resource managers than a nominal gradient of change classes, which we have used for previous project areas. The classifier we are testing for this project was developed by Xueqiao Huang and John Jensen as a module integrated within the ERDAS IMAGINE image processing environment (Huang and Jensen, 1997). The classifier uses the inductive learning algorithm C4.5 (Quinlan, 1993) to automate the building of a knowledge base related to changes in vegetation. A decision tree strategy is employed to efficiently generate production rules from multiple layers of training data.

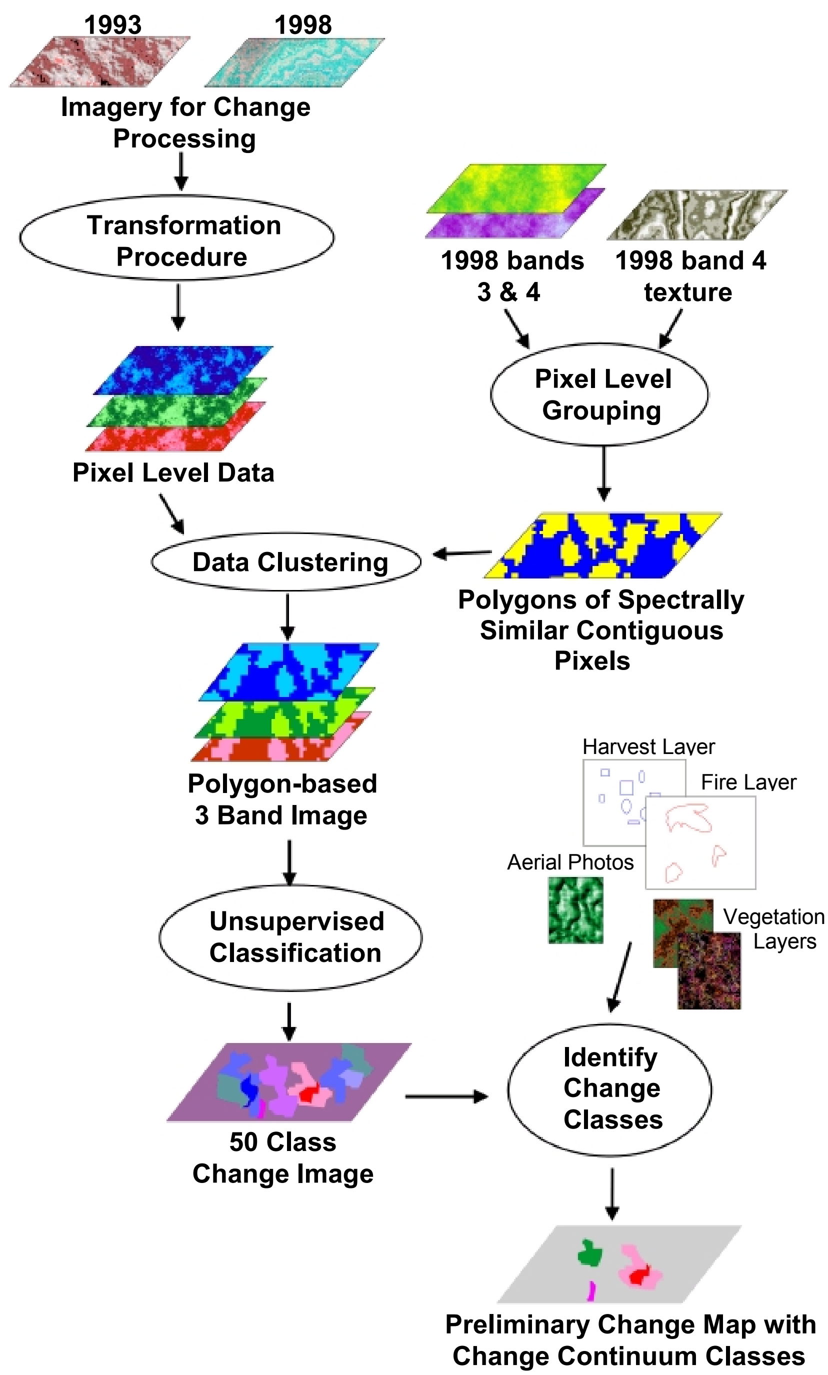

Figure 3. Phase I Methods.

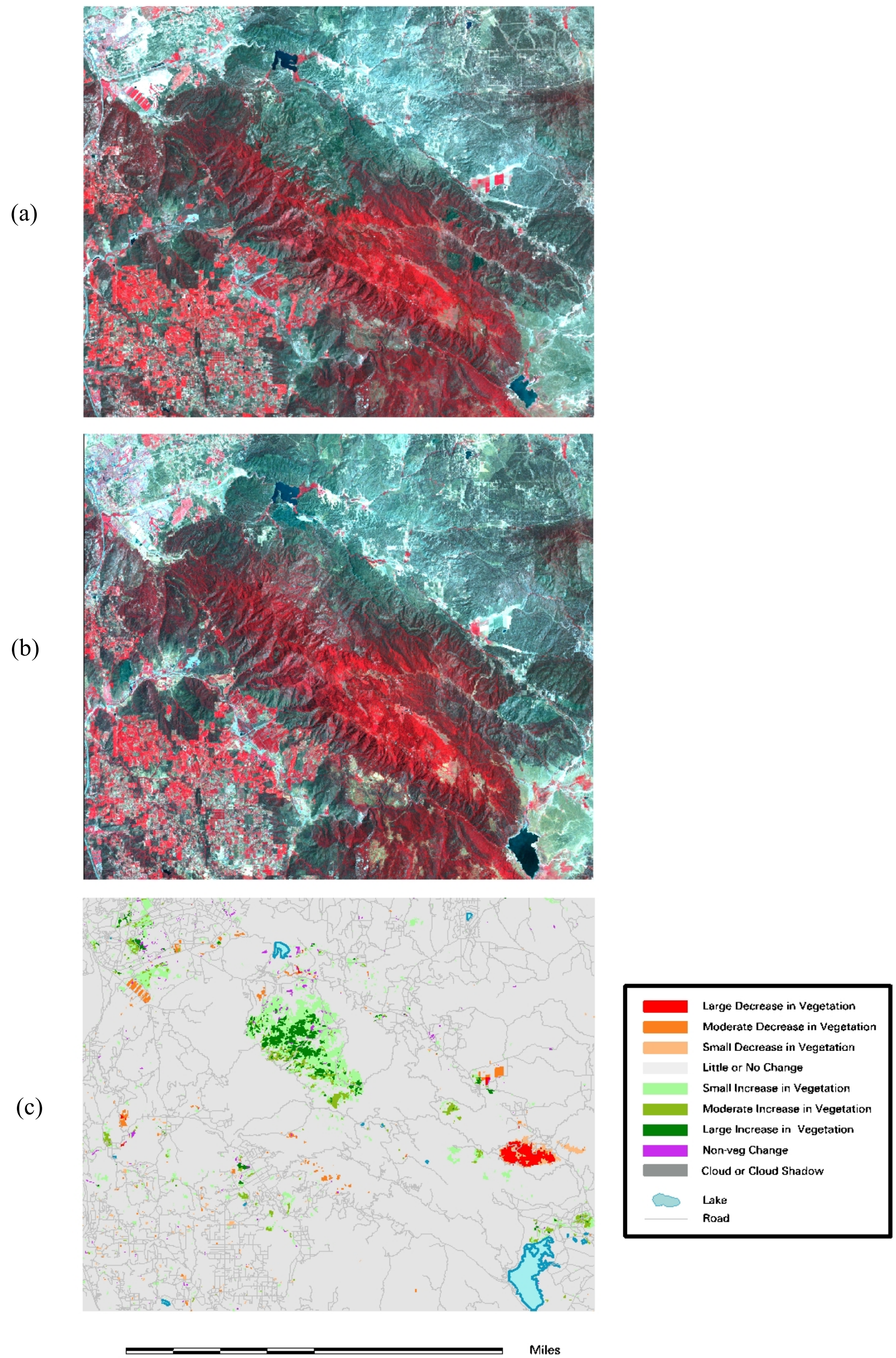

Figure 4. Landsat TM imagery and Phase I change data for the pilot area in the southern California project area. (a) 1990 TM image, (b) 1996 TM image, (c) change data.

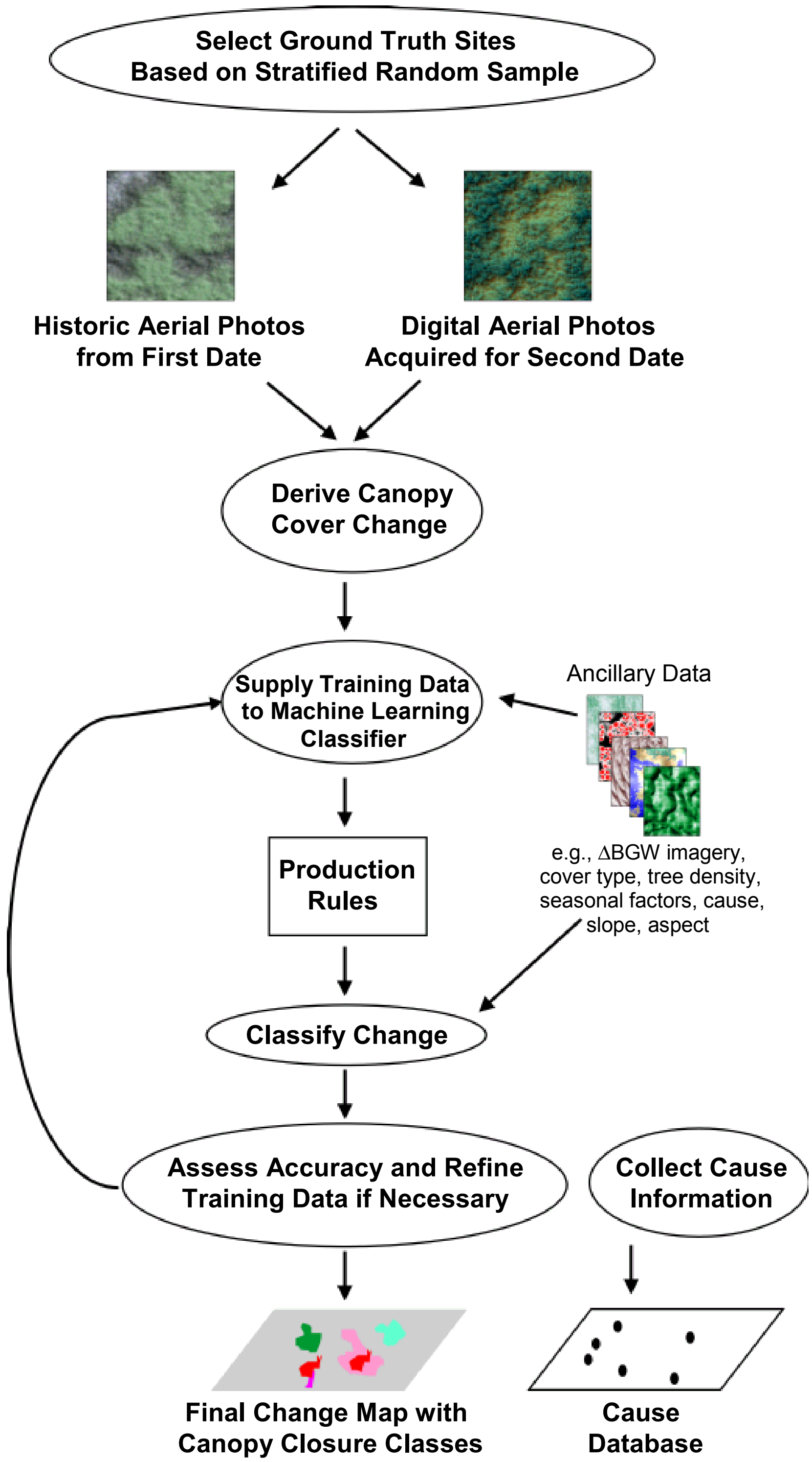

Figure 5. Phase II Methods.

TABLE 1. CLASSIFICATION SCHEMES FOR PHASE I AND PHASE II CHANGE MAPS

PHASE I CHANGE CLASSES	PHASE II CHANGE CLASSES
Large Decrease in Vegetation	–71 to –100 % CC (canopy cover)
Moderate Decrease in Vegetation	–41 to –70 % CC
Small Decrease in Vegetation	–16 to –40 % CC
Little or No Change	+15 to –15 % CC (Little or No Change)
Small Increase in Vegetation	+16 to +40 % CC
Moderate Increase in Vegetation	+41 to +100 % CC
Large Increase in Vegetation	Shrub or Grass Decrease > 15 %
Non-Vegetation Change	Shrub or Grass Increase > 15 %
Terrain Shadow	Vegetation Decrease within Existing Urban Area
Cloud or Cloud Shadow	Vegetation Increase within Existing Urban Area
	Terrain Shadow
	Cloud or Cloud Shadow

As with a neural network, there are several advantages to using a machine-learning approach. Since ancillary data layers may be used to help improve discrimination between classes, fewer field samples are generally required for training. This machine learning model is non-parametric and does not require normally-distributed data or independence of attributes. It can also recognize nonlinear patterns in the input data that are too complex for conventional statistical analyses or too subtle to be noticed by an analyst. Once suitable production rules have been developed for a given project area, they may be quickly applied the next time that area is revisited without having to retrain the classifier. Areas of similar landcover may require only a slight modification of the rules. This could represent a considerable time savings during the classification stage of our future project areas, in addition to providing us with an empirical understanding of the relationships between our variables.

Training layers for the machine-learning classifier are assembled using data from field sites and other ancillary data. Layers currently being tested with the classifier include the interpreted canopy cover change (ground truth), transformed change data, covertype, tree density, harvest/fire history, local climate, slope, aspect, and time interval between acquisition dates.

Ground truth is obtained by estimating canopy cover change within a five-year time frame using two sets of aerial photographs and field site information. Ground truth sites are selected using a stratified random sampling scheme based on change/lifeform classes. Aerial imagery for the first date is provided by either 1:15,840-scale color-IR resource photography or digital orthoquads with a one-meter ground resolution cell. A DCS420 Kodak digital color infrared camera is used to collect aerial imagery for the second date. In order to calibrate canopy cover estimates from the photos, a subsample of ground truth sites is selected to visit in the field, representing approximately one quarter of the 284 digital camera sites. Canopy cover measurements are recorded at each site on the ground, and attributes such as species and ground cover were noted. Dot grid measurements of canopy cover are then gathered from the aerial imagery for each date. Approximately half of the photo sites are reserved for accuracy assessment.

The training data are compiled as a layered stack of images, and supplied to the classifier in the form of a text file with each line representing a training object (or pixel) with its associated attribute/class information. The classifier makes use of this information within its learning subsystem to create a knowledge base of production rules. The knowledge base is then used by the expert classifier subsystem to produce a change image portraying classes of crown closure change. The new change map is assessed for accuracy using the photo sites set aside earlier. The production rules may be used to help analyze the nature of any misclassifications in the map. If necessary, revisions can be made to the training data to better represent the conditions present, and the classifier run again until a suitable level of accuracy has been achieved. Before the final change map is ready for distribution, changes occurring within the delineated boundaries of clouds or cloud shadows are recoded to the “Cloud or Cloud Shadow” class.

The final step in Phase II is to identify causes of change and create a database of points with associated causal information. This process begins by overlaying fire, harvest, and plantation layers onto the change detection map in a

GIS. This process readily attributes areas of change due to wildfire, prescribed fire, management practices, and vegetation regrowth. The change map and imagery are also used to interpret and delineate areas of urban development within the project area. Once all known causes have been identified, 7.5-minute quadrangle-size change maps are created for the unlabeled areas of change in conifer and hardwood vegetation types. Base layers from USGS digital raster graphic (DRG) quadrangles (such as topography, roads, and annotation) are used on the change quads to provide a familiar frame of reference for map users. National Forest resource specialists interpret the conifer change maps by applying local knowledge regarding sources of change in coniferous forests. Similarly, resource specialists from the University of California Integrated Hardwood Rangeland Management Program (IHRMP) consult private landowners to identify sources of change in hardwood rangelands.

Collecting field data on National Forest's and hardwood rangelands further aids in interpreting natural and human-induced change. Fieldwork conducted by IHRMP resource specialists in the hardwood rangelands has identified causes of changes in canopy cover due to fire, thinning, harvest, urban development, mortality, regeneration, and tree planting. Areas of mortality, recent fires, and timber harvest not included in our current ancillary data were identified on National Forest's within the project areas.

The final product from Phase I is a change map containing a gradient of classes that range from large decreases in vegetation to large increases in vegetation. Phase II products include the enhanced change map from the machine learning classifier, featuring discrete canopy cover change classes (Table 1, right side), and the GIS database identifying the locations of vegetation change with cause information for coniferous forestland, hardwood rangeland, shrub landcover, and urban areas.

RESULTS

At press time, processing with the machine-learning classifier was underway for the southern California project area. Results should be available at the May 1999 Portland ASPRS conference. Preliminary results for a pilot area in Lassen National Forest (in the northeastern California project area) indicate that this method has the potential to reveal more subtle changes in vegetation than our traditional Phase I change map. For example, the machine-learning classifier output indicated small increases in vegetation within numerous plantations that were labeled as "Little to No Change" in our original Phase I output. The increases in vegetation were verified using resource photography and tabular plantation data.

SUMMARY

The USDA Forest Service, in collaboration with the California Department of Forestry and Fire Protection, is conducting a statewide change detection program to provide landcover change across all ownerships. We are currently testing a new methodology for production of change maps based on discrete canopy cover classes rather than a nominal gradient of change categories. The new technique uses a machine-learning classifier to categorize change and digital airphotos to collect ground truth.

The machine-learning classifier uses an inductive learning algorithm to automate the building of a knowledge base related to changes in vegetation. A decision tree strategy is employed to generate production rules from multiple layers of training data. The knowledge base is then used by the expert classifier subsystem to produce a change image based on changes in crown closure. This approach provides a way to incorporate into the classification other sources of information related to vegetation change, and learn more about the relationships between these data layers through the generated knowledge base.

Ground truth for the classifier is obtained by estimating canopy cover change over the five-year timeframe using color-IR digital photos, digital orthophoto quads, and aerial photography. This strategy is intended to replace our previous method of collecting detailed inventory data, which was time-intensive and cost-prohibitive over such large project areas, and produced an insufficient number of field sites for correlation with detected changes in vegetation. Should the new technique prove feasible, it will provide a much larger number of ground truth sites for use in classification and accuracy assessment, at a lower cost. The main challenge has been finding a suitable source of imagery for the first image date. This difficulty should be alleviated once we come full-circle on our five-year revisit cycle. Preliminary results indicate that the machine-learning classifier may help increase the utility and accuracy of our traditional change map product.

REFERENCES

Collins, J.B., and C.E. Woodcock, 1995. Assessment of drought-induced conifer mortality in the Lake Tahoe Basin using remotely sensed imagery, *Technical Report No. 11*, USDA Forest Service, Region 5.

Collins, J.B., and C.E. Woodcock, 1996. An assessment of several linear change detection techniques for mapping forest mortality using Multitemporal Landsat TM data, *Remote Sensing of Environment,* 56(1): 66-67.

Huang, X., and J. Jensen, 1997. A Machine-Learning Approach to Automated Knowledge-Base Building for Remote Sensing Image Analysis with GIS Data, *Photogrammetric Engineering & Remote Sensing*, 63(10):1185-1194.

Macomber, S., and C.E. Woodcock, 1994. Mapping and monitoring conifer mortality using using remote sensing in the Lake Tahoe basin, *Remote Sensing of Environment*, 50:255-266.

Quinlan, J.R., 1993. *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, San Mateo, California.

Ryherd, S.L., and C.E. Woodcock, 1990. The use of texture in image segmentation for the definition of forest stand boundaries, Presented at the Twenty-Third International Symposium on Remote Sensing of Environment, Bangkok, Thailand, April 18-25.

Tanre, D., C. Deroo, P. Duhaut, M. Herman, J.J. Morcrette, J. Perbos, and P.Y. Deschamps, 1990. Description of a computer code to simulate the satellite signal in the solar spectrum: the 5S code*, International Journal of Remote Sensing*, 11:659-668.

Woodcock, C.E., J.B. Collins, S. Gopal, V. Jakabhazy, X. Li, S. Macomber, S. Ryherd, V.J. Harward, J. Levitan, Y. Wu, and R. Warbington, 1994. Mapping forest vegetation using Landsat TM imagery and a canopy reflectance model, *Remote Sensing of Environment*, 50:240-254.