

Interactive Training of Advanced Classifiers for Mining Remote Sensing Image Archives

Selim Aksoy
Bilkent University
Department of Computer Engineering
Bilkent, 06800, Ankara, Turkey
saksoy@cs.bilkent.edu.tr

Krzysztof Koperski, Carsten Tusk,
Giovanni Marchisio
Insightful Corporation
1700 Westlake Ave. N., Suite 500
Seattle, WA, 98109, USA
{krisk,ctusk,giovanni}@insightful.com

ABSTRACT

Advances in satellite technology and availability of downloaded images constantly increase the sizes of remote sensing image archives. Automatic content extraction, classification and content-based retrieval have become highly desired goals for the development of intelligent remote sensing databases. The common approach for mining these databases uses rules created by analysts. However, incorporating GIS information and human expert knowledge with digital image processing improves remote sensing image analysis. We developed a system that uses decision tree classifiers for interactive learning of land cover models and mining of image archives. Decision trees provide a promising solution for this problem because they can operate on both numerical (continuous) and categorical (discrete) data sources, and they do not require any assumptions about neither the distributions nor the independence of attribute values. This is especially important for the fusion of measurements from different sources like spectral data, DEM data and other ancillary GIS data. Furthermore, using surrogate splits provides the capability of dealing with missing data during both training and classification, and enables handling instrument malfunctions or the cases where one or more measurements do not exist for some locations. Quantitative and qualitative performance evaluation showed that decision trees provide powerful tools for modeling both pixel and region contents of images and mining of remote sensing image archives.

Categories and Subject Descriptors: I.5.2 [Pattern Recognition]: Design Methodology; I.4.8 [Image Processing and Computer Vision]: Scene Analysis

General Terms: Algorithms, Design, Experimentation

Keywords: Decision tree classifiers, missing data, data fusion, remote sensing, land cover analysis

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

KDD'04, August 22–25, 2004, Seattle, Washington, USA.
Copyright 2004 ACM 1-58113-888-1/04/0008 ...\$5.00.

1. INTRODUCTION

Remotely sensed imagery has become an invaluable tool for scientists, governments, military and the general public to understand the world and its surrounding environment. Military applications include using land cover classification to derive tactical decision aids like landing zone and troop movement plans, to manage ecosystems of military lands and waterways, and to study interactions between mission activities and ecological patterns. Similarly, land cover maps help civil users to perform land use monitoring and management, fire protection, site suitability, agricultural and public health studies.

The amount of data received from satellites is constantly increasing. For example, NASA's Terra satellite sends more than 850GB of data to Earth every day [15]. Automatic content extraction, classification and content-based retrieval have become highly desired goals to develop intelligent remote sensing image databases. State-of-the-art remote sensing image analysis systems aid users by providing classification tools that use spectral information or texture features as the input for statistical classifiers that are built using unsupervised or supervised algorithms. The most commonly used classifiers are the minimum distance classifier and the maximum likelihood classifier with a Gaussian density assumption. Spectral signatures and other features do not always map conceptually similar patterns to nearby locations in the feature space and limit the success of minimum distance classifiers. Furthermore, these features do not always have Gaussian distributions so maximum likelihood classifiers with this assumption fail to model the data adequately.

Like any data analysis problem, domain knowledge and prior information is very useful in land cover classification. Incorporating supplemental GIS information and human expert knowledge into digital image processing has long been acknowledged as a necessity for improving remote sensing image analysis [8]. Artificial intelligence research and developments in rule-based classification systems have enabled a computer to mimic the heuristics and knowledge that a human expert uses in interpreting an image. Consequently, rule-based classification systems [10] have been successfully used for fusion of spectral and ancillary data in applications such as land cover classification [8, 11], land cover change monitoring [17], sea ice classification [18], and ridge line extraction from Digital Elevation Model (DEM) data [14].

However, a common problem in all of these attempts has been the translation of expert knowledge to a computer-

usable format. First, remotely sensed data are converted into formats supported by statistical analysis packages; then, these packages are used to create rules; finally, the rules are re-formatted so that they can be used in a rule-based image classifier [8, 11]. Today, commercial-of-the-shelf remote sensing image analysis systems have rule-based classification modules but they operate on individual scenes and require an expert to create the rules. This requirement for an enormous amount of manual processing even for small data sets makes knowledge discovery in large remote sensing archives practically impossible.

The VISIMINE system [9] we have developed supports interactive classification and retrieval of remote sensing images by modeling them on pixel, region and scene levels. Pixel level characterization provides classification details for each pixel with regard to its spectral, textural and other ancillary attributes. Following a segmentation process, region level features describe properties shared by groups of pixels. Scene level features model the spatial relationships of the regions composing a scene using a visual grammar. This hierarchical scene modeling with a visual grammar aims to bridge the gap between features and semantic interpretation.

This paper describes our work on developing decision tree-based tools to automate the process of acquiring knowledge for image analysis and mining in remote sensing archives. Decision tree classifiers are used to model image content in pixel and region levels. VISIMINE provides an interactive environment for training customized semantic land cover labels from a fusion of visual attributes. Spectral values, ancillary GIS layers and other image attributes like textural features are used in decision tree learning. The learning process also includes algorithms for handling missing data and automatic conversion of models into rule bases.

The rest of the paper is organized as follows. The system architecture and data used in the experiments is introduced in Section 2. The classifiers used for land cover modeling are described in Section 3. User interaction tools for training and evaluating these models are presented in Section 4. Performance evaluation for the proposed system is discussed in Section 5. Conclusions are given in Section 6.

2. PROTOTYPE SYSTEM DESIGN

VISIMINE system architecture is given in Figure 1. The image processing, statistics, machine learning and data mining libraries are implemented in C/C++. The graphical user interface is implemented in Java. Oracle is used as the RDBMS to store and access images, features and metadata. VISIMINE also has a seamless interface with Insightful's S-PLUS statistical analysis software.

The input to the system is raw image and ancillary data. This data is automatically processed by unsupervised algorithms in the image processing library for feature extraction. Original data and extracted features become the input to classification algorithms in the machine learning library. The user interacts with the system by providing a list of land cover labels and corresponding training examples. Then, the machine learning library seamlessly transfers this expert knowledge into classification models in terms of decision trees and rules. Intuitive but powerful visualization tools allow the user examine the classification models and provide feedback to the system. These models can be used to classify other images in the same data set, or can be used to search other collections for similar scene structures.

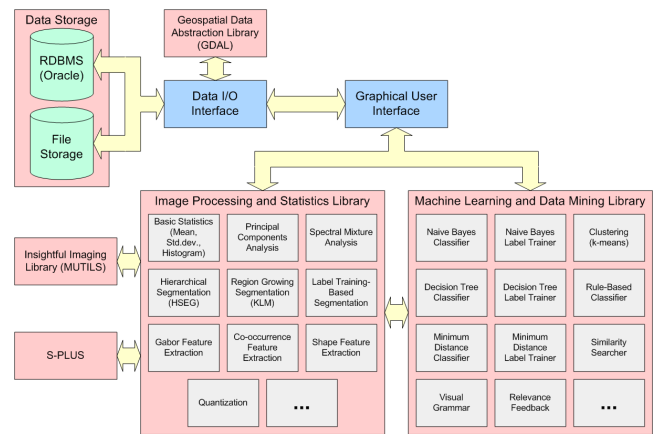


Figure 1: VISIMINE system architecture. The image processing library contains unsupervised algorithms to process both spectral and ancillary data to do feature extraction. The machine learning library contains automatic and semi-automatic algorithms to train higher-level class models. User can interact with the system to provide training examples and feedback to update the models.

2.1 Pixel Level Data

The image and ancillary raster data used for evaluating the system described in the paper consist of Aerial (1.8m resolution, 3 bands), Ikonos (4m resolution, 4 bands, 2 sets) and DEM (30m resolution, 1 band) data layers that cover the Fort A.P. Hill area in Virginia and are provided by the U.S. Army Topographic Engineering Center. These layers were upsampled to the same resolution (1.8m) where each band has $11,683 \times 11,677$ pixels.

We also extracted Gabor wavelet features [6, 13] for micro-texture analysis on the first band of Aerial data. (We also extracted Gabor features using the other Aerial and Ikonos bands but did not observe any significant improvement in the results.) Gabor features were computed by filtering an image with Gabor wavelet kernels at different scales and orientations. We used kernels rotated by $n\pi/8$, $n = 0, \dots, 7$, at two scales. To obtain rotation invariant features, we computed the autocorrelation of the wavelet filter outputs with 0 and 90 degree phase differences at each scale. This resulted in four bands corresponding to two phase differences for each of the two scales.

The total size of the data is about 4.5GB. False color composite images of some of the data layers and their coverages are shown in Figure 2. The ground truth includes 11 pixel level land cover labels with independent training and testing data described in Table 1.

2.2 Region Level Data

The pixel level classification results can be converted to spatially contiguous region representations using a segmentation algorithm [1] that iteratively merges pixels with identical labels as illustrated in Figure 3. The region segmentation algorithm produces region polygon boundaries that can be used to extract features to build region level land cover labels. These region level features include the following shape features [7, 5] computed directly from the region boundaries:

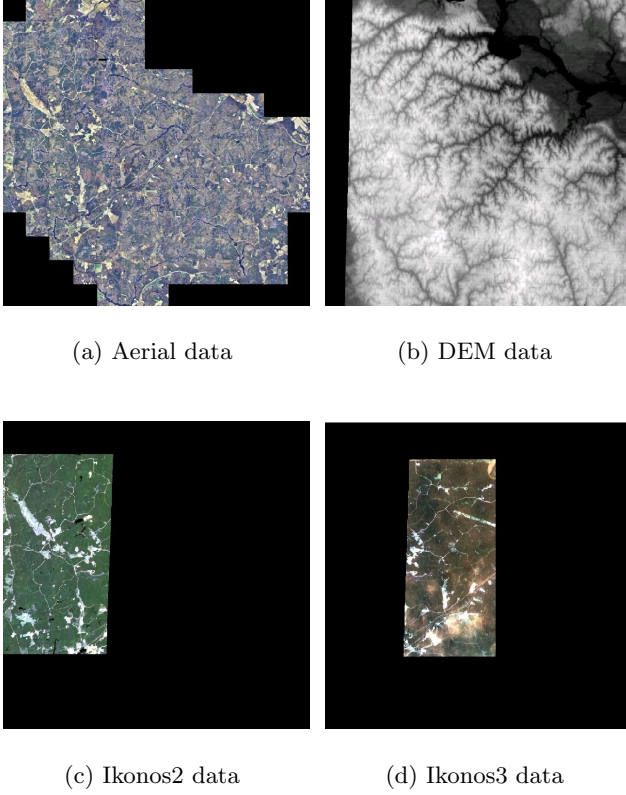


Figure 2: Data for the Fort A.P. Hill scene used in the experiments.

1. Bounding box coordinates (MINX, MAXX, MINY, MAXY),
2. Area (AREA),
3. Perimeter (PERIMETER),
4. Centroid (center of mass) on x and y axes (CENTROIDX, CENTROIDY),
5. Spatial variances on x and y axes (SIGMA_X, SIGMA_Y),
6. Spatial variances on region's principal (major and minor) axes (SIGMA_MAJOR, SIGMA_MINOR),
7. Orientation of the major axis with respect to the x axis (THETA),
8. Eccentricity (ratio of the distance between the foci to the length of the major axis; e.g. a circle is an ellipse with zero eccentricity) (ECCENTRICITY),
9. Normalized central moments (central moments are computed by shifting the origin to the centroid and normalized central moments are computed by dividing the central moments by normalization factors that are functions of the moments' orders) (NORM_CENT_11, NORM_CENT_20, NORM_CENT_02, NORM_CENT_21, NORM_CENT_12),

Table 1: Land cover labels and the number of training and testing examples used in the experiments. Each label is represented by the corresponding color in the figures in the rest of the paper.

Land cover	Color	# training examples	# testing examples
burned		145	456
paved		91	536
building		664	427
ground		2,434	2,442
crop		9,433	49,765
grass		7,223	15,329
brush		3,117	7,170
pine		11,284	60,669
deciduous		10,409	59,511
water		14,275	41,502
marsh/wetland		7,204	8,097
Total		66,279	245,904

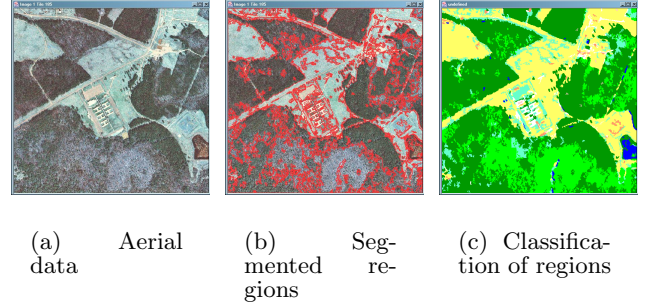


Figure 3: Segmenting regions using pixel level classification. Connected pixels having the same labels in the pixel level classification map are iteratively merged using area constraints. Resulting segmentation boundaries are shown in red in the second figure.

10. Invariant moments (computed as combinations of normalized central moments to have invariance to translation, rotation and scale change) (INVARIANT_1, INVARIANT_2, INVARIANT_3, INVARIANT_4, INVARIANT_5, INVARIANT_6, INVARIANT_7).

In addition to these 26 shape features listed above, each region is also assigned a land cover label (LABELID) propagated from its pixels (land cover label of a region is the common land cover label of its pixels). We also used statistical summaries of the pixel contents of regions as region level features. These summaries include means, standard deviations and histograms of pixel level features (e.g. DEM, Aerial, Gabor) within individual regions.

3. LABEL TRAINING

We define label training as the process of using machine learning algorithms to model high-level user-defined semantic land cover labels in terms of low-level image and ancillary data attributes. We used decision tree models and rule-

based models to enable a computer to mimic the heuristics and knowledge that a human expert uses in interpreting an image. Details of decision tree and rule-based models are given next.

3.1 Decision Tree Classifier

Decision trees are used to predict a categorical response (class) based on a collection of predictors (attributes). Tree-based tools offer the following advantages:

- They can operate on both numerical (continuous) and categorical (discrete) measurements.
- They do not require any assumptions about neither the distributions nor the independence of attribute values. This is especially important for the fusion of measurements from different sources like spectral data, DEM data and other ancillary GIS data.
- They are often easy to interpret, even by those with no statistical expertise, by creating subgroups of data which the user may graphically analyze.
- They are capable of dealing with missing data during both training and classification.

We build decision trees using the RPART (recursive partitioning) learning algorithm [19]. RPART uses Gini and entropy split functions to partition the data. It searches through the attributes one by one and for each attribute finds the best split. Then, it compares the best single attribute splits and selects the best of the best. Next, the data is separated into two using that split, and this process is recursively applied to each sub-group until the sub-groups either reach a minimum size or until no improvement can be made. Once the leaf nodes are found, they are labeled by the class that has the most objects represented. The confidence value for that class is computed as the ratio of training objects that belong to that class to the total number of objects in that node. A cost-complexity measure with cross-validation is used to avoid very extensive trees for increased generalization ability.

An important problem in statistical modeling and analysis is the presence of missing data. There are several possible reasons for a value to be missing, such as: it was not measured; there was an instrument malfunction; the attribute does not apply, or the attribute's value cannot be known. This is also an important problem in remote sensing image analysis where one or more data bands may be completely missing or there may be gaps in coverage for particular regions at particular times (as can be seen in the coverages of our test data in Figure 2) because of satellite orbit restrictions, heavy clouds, haze or other atmospheric conditions, and viewing and illumination geometry.

Most algorithms "deal" with missing data by ignoring objects with incomplete measurements. This is quite wasteful because remote sensing data is often hard and expensive to obtain. A common technique is to calculate impurities using only the attribute information present so that any object with at least one attribute will participate in training. Another ad hoc solution is to replace a missing attribute by its mean or a randomly generated value using a parametric model estimated using the training objects that are not missing that attribute. However, this biases and distorts the marginal distributions of the attributes [12].

We handle missing data using surrogate splits. The idea behind surrogate splits is to use the primary decision attribute at a node whenever possible, and use alternative attributes when the object is missing that attribute. This can be achieved by an ordered set of surrogate splits for each non-leaf node [3, 4]. The first surrogate split maximizes the probability of making the same decision as the primary split, i.e. the number of objects that are sent to the same descendant branches by both the primary and the surrogate splits is as high as possible. Other surrogate splits are defined similarly. If an object is missing all the surrogate attributes, the blind rule is used, i.e. it is sent to the descendant node that received most of the training objects.

3.2 Conversion of Trees into Rules

At any time of the learning process, decision trees can be automatically converted to decision rules. This can be done by tracing the tree from the root node to each leaf node and forming logical expressions that make the initial set of rules. Occasionally, some of these rules can be redundant and can be simplified without affecting the classification accuracy. We investigated the following schemes for rule generalization:

- Lossless generalization where conditions that are completely redundant with respect to other conditions are removed.
- Lossy generalization where conditions are removed using greedy elimination. This is done by comparing error estimates of the original rule and the resulting rule with one of the conditions deleted. If the error rate for the latter case is no higher than that of the original rule, that condition is deleted. We used the pessimistic error estimate [16] where, given a confidence level, the upper limit on the probability of error is computed using the confidence limits for the Binomial distribution.

We also further simplify the rules by deleting the ones that have error estimates that are greater than the error estimate for the default rule. The default rule is used to assign the observations that do not satisfy any rule to the class with the highest frequency in the training data.

Examples of rule generalization are given in Figure 4. Decision trees always give mutually exclusive sets of rules. However, rules may not stay mutually exclusive after the rule generalization process. To avoid conflicts, we sort the rules in descending order of the probability values. If an observation satisfies none of the rules, it is assigned to the default class that appears most frequently in the training set.

3.3 Rule-based Classifier

We developed a rule-based classifier that uses the simplified rule set to classify an image. This classifier also supports surrogate splits in the conditions and can handle missing data. The rules are checked in descending order of the probability (confidence) values. A rule is satisfied if all of its conditions are satisfied. The default rule is used to assign the observations that do not satisfy any rule to the class with the highest frequency in the training data.

```

non-generalized rule:
  IF AERIAL_GABOR::FINEODEG >= 66.3421
    AND AERIAL_GABOR::COARSEODEG < 253.842
    AND AERIAL::BAND1 < 142.5
    AND AERIAL::BAND2 < 76.5
    AND DEM::ELEVATION < 50.5
    AND DEM::ELEVATION < 10
    AND DEM::ELEVATION < 5.5
  THEN CLASS water WITH PROB 1
generalized rule:
  IF DEM::ELEVATION < 5.5
  THEN CLASS water WITH PROB 0.99923

non-generalized rule:
  IF AERIAL_GABOR::FINEODEG >= 66.3421
    AND AERIAL_GABOR::COARSEODEG >= 253.842
    AND AERIAL::CLUSTERID in {10,11,13,15,16,17,18,19,20,21,22}
    AND DEM::ELEVATION >= 45.5
    AND AERIAL_GABOR::COARSEODEG >= 476.393
    AND DEM::ELEVATION < 60.5
  THEN CLASS deciduous WITH PROB 0.666667
generalized rule:
  IF AERIAL_GABOR::COARSEODEG >= 476.393
    AND DEM::ELEVATION < 60.5
  THEN CLASS deciduous WITH PROB 0.85664

```

- removed based on error estimate
- removed based on error estimate
- removed based on error estimate
- removed based on error estimate
- redundant w.r.t. ELEVATION < 5.5
- redundant w.r.t. ELEVATION < 5.5

- removed based on error estimate
- redundant w.r.t. COARSEODEG >= 476.393
- removed based on error estimate
- removed based on error estimate

Figure 4: Rule generalization examples.

4. USER INTERACTION

Our goal in designing the graphical user interface (GUI) has been to develop an environment with an interactive learning component that requires minimal user interaction. The design constraints included minimizing the amount of interaction required and making the work flow from raw data to the final classification as transparent as possible to the user.

4.1 Label Training

The GUI for label training is shown in Figure 5. The only interaction required is user's providing of examples for land cover labels. These examples are input to the system using mouse clicks. The user can view the color composite image or individual data bands while entering examples in the training display. The GUI allows users to add both training and testing (ground truth) examples.

Each user can create a new model or update models previously created by other users. At any time of the learning process, the user can validate the current classification model with the ground truth data using the automatically generated confusion matrices. Advanced users can also change the parameters of the classification models through the GUI. The models can be stored either in the RDBMS or in a file.

4.2 Tree and Rule Visualization

The GUI for tree visualization is shown in Figure 6. Each node in the tree shows the primary split condition and, for each label, percentages of training examples that satisfy that condition. Prediction importance of each data layer is measured as the total reduction in the split criterion achieved by that data layer. Two Aerial bands, Gabor texture features and DEM were found to be the most important predictors in our experiments. Details of a node selected in the tree display are also shown for further examination by an advanced user. These details include the number of training examples



Figure 5: GUI for label training. The panel on top-left shows the land cover labels defined by the user. The user can add new labels or remove existing ones, and can also change the color of a label. The panel on bottom-left shows the data layers used in training. The user can add or remove layers using the corresponding buttons. Double-clicking on a data layer shows that layer in the left image panel. The “load tile” button loads a new image for training and/or classification. The “update” button shows the probability map for an individual label or the classification map for selected labels in the right image panel. The “show info” button displays information about the trained classifier. The “show tree” and “show rules” buttons open the tree and rule displays respectively.

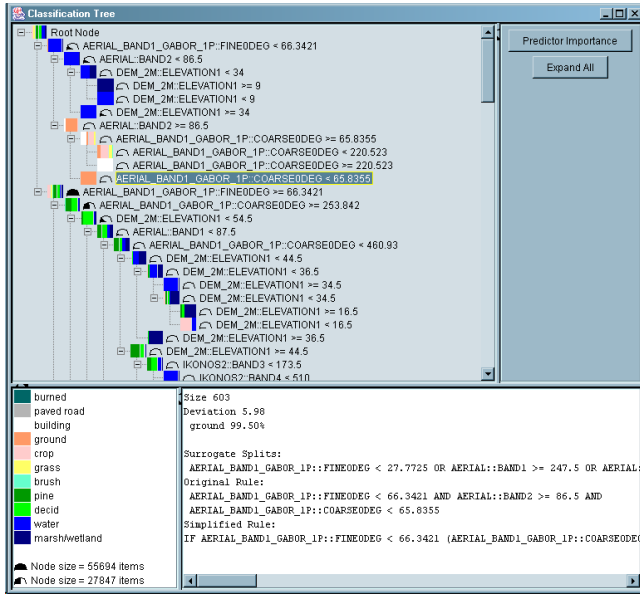


Figure 6: GUI for tree visualization.

(size), the Gini or entropy impurity value (deviation), surrogate splits, path from the root node to the current node (original rule), and the generalized rule if it is a leaf node.

4.3 Tree and Rule Tracing

In tree tracing, the user can select a node in the tree and see which pixels/regions pass through that node during classification. In rule tracing, the user can select a rule and see which pixels/regions are classified using that rule. Tracing also allows the user to select a pixel/region in the original image and see which node and rule are used to classify that pixel/region.

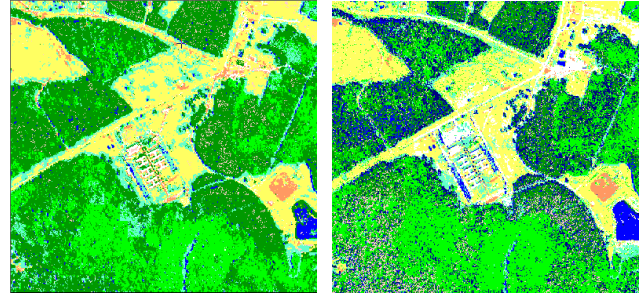
5. PERFORMANCE EVALUATION

We evaluated the performance of the system using the four data layers (consisting of a total of 16 bands) described in Section 2. We present quantitative and qualitative results for pixel level and region level classification below.

5.1 Pixel Level Classification

5.1.1 Evaluation of classification with different data layers

Examples of using different data layers for classification are given in Figure 7. The corresponding confusion matrices are given in Figures 8(a) and 8(b). Using all 16 bands resulted in a high classification accuracy of 72%. When we used only the bands with prediction importance higher than 2000 (7 bands), classification accuracy became slightly higher, because of better pruning and better classification of crop fields. Without the DEM band classification accuracies for crop and brush dropped drastically. When we used only the visual bands (without the Gabor texture features and the DEM band), classification accuracy dropped close to 50%. Note that the confusion matrices show zero results for recognition of paved roads and burned areas. This is due to lack of qualified ground truth data as visual evaluation acknowledges correct classification of roads in many cases



(a) Using all 16 bands

(b) Using 11 visual bands

Figure 7: Classification using all 16 bands from Aerial, Gabor, DEM, Ikonos versus using 11 visual bands from Aerial and Ikonos. We observe a good overall classification when all bands are used. Without texture and DEM information more forest areas are misclassified as marsh or crop because the effects of shadows in noisy visual bands become more apparent in the absence of texture.

(see Section 5.2). Finally, we experimented with different parameter settings for the decision tree classifier. The differences caused by using different split functions were very subtle in our dataset.

5.1.2 Evaluation of rule generalization

The confusion matrix for classification using generalized rules is given in Figure 8(c). Classification accuracy for generalized rules was very similar to (even slightly higher than) the one for the decision tree classifier. The improvement was due to the additional pruning during rule generalization.

5.1.3 Evaluation of classification with different classifiers

First, we compared decision tree classifiers with the minimum distance classifier (confusion matrix is given in Figure 8(d)). The minimum distance classifier cannot handle missing data so we used only three Aerial and four Gabor bands for both classifiers. The decision tree classifier appeared superior, not only in its capability to handle numerical as well as categorical data and to cope with missing data values, but also in regard to handling pure numerical feature data. The comparison showed an overall average classification accuracy of 48% using the minimum distance classifier compared to 64% using the decision tree classifier. Both experiments were performed using the same training and validation data for both classifiers.

We also compared decision tree classifiers with the unsupervised k -means clustering algorithm. We ran k -means using 11 clusters with three Aerial bands. Example classifications using the same three bands with the decision tree classifier, minimum distance classifier and k -means clustering are given in Figure 9. As expected, unsupervised clustering resulted in the most noisy results. Some of the building and grass pixels were classified correctly but most of the other labels were not separated successfully. Absence of Gabor features also caused the decision tree and minimum

Classification Agreement

		Assigned											Total	% Agree
		burned	paved road	building	ground	crop	grass	brush	pine	decid	water	unclassified		
True	burned	0	0	0	0	0	0	68	117	8	0	263	456	0.00%
	paved road	0	0	55	71	114	154	97	13	23	9	0	536	0.00%
	building	0	0	307	4	9	40	45	3	18	0	1	427	71.90%
	ground	0	0	26	1569	356	147	264	47	33	0	0	2442	64.25%
	crop	0	0	26	1801	31182	4927	429	4178	448	175	6599	49765	62.66%
	grass	0	0	19	457	1326	9611	3279	409	119	109	0	15329	62.70%
	brush	0	0	139	243	263	1512	4233	449	306	25	0	7170	59.04%
	pine	0	0	37	0	2888	42	799	51085	2532	1389	1897	60669	84.20%
	decid	0	0	209	21	145	102	4991	7048	44488	405	2102	59511	74.76%
	water	0	0	0	18	119	11	2	570	160	30682	9940	41502	73.93%
	unclassified	0	0	0	0	154	43	0	1146	1800	295	4659	8097	57.54%
Total		0	0	818	4184	36556	16589	14139	65016	50044	33097	25461	245904	72.31%

(a) Using decision tree with all 16 bands from Aerial, Gabor, DEM and Ikonos

Classification Agreement

		Assigned											Total	% Agree
		burned	paved road	building	ground	crop	grass	brush	pine	decid	water	unclassified		
True	burned	0	0	0	0	0	0	0	36	148	9	263	456	0.00%
	paved road	0	0	35	1	114	187	143	13	26	17	0	536	0.00%
	building	0	0	308	3	9	34	1	3	24	44	1	427	72.13%
	ground	0	0	94	1500	356	263	63	47	33	86	0	2442	61.43%
	crop	0	0	27	812	34152	894	1676	4030	1316	244	6614	49765	68.63%
	grass	0	0	27	1004	1454	9072	2655	398	220	499	0	15329	59.18%
	brush	0	0	144	2	282	1268	4442	441	360	231	0	7170	61.95%
	pine	0	0	21	2	2686	41	447	50701	3270	1616	1885	60669	83.57%
	decid	0	0	211	0	248	81	1491	6931	45286	3161	2102	59511	76.10%
	water	0	0	0	18	120	0	0	520	223	30681	9940	41502	73.93%
	unclassified	0	0	0	0	155	0	0	1079	1910	294	4659	8097	57.54%
Total		0	0	867	3342	39576	11840	10918	64199	52816	36882	25464	245904	73.53%

(b) Using decision tree with 7 bands (two Aerial, two Gabor, one DEM and two Ikonos) with prediction importance over 2000

Classification Agreement

		Assigned											Total	% Agree
		burned	paved road	building	ground	crop	grass	brush	pine	decid	water	unclassified		
True	burned	0	0	0	0	0	0	36	148	12		260	456	0.00%
	paved road	0	0	38	2	74	206	119	10	67	20	0	536	0.00%
	building	0	0	302	9	3	34	1	0	34	43	1	427	70.73%
	ground	0	0	95	1820	2	301	61	8	41	114	0	2442	74.53%
	crop	0	0	31	832	33916	927	1645	4122	1228	390	6674	49765	68.15%
	grass	0	0	30	1003	1321	9348	2339	401	261	626	0	15329	60.98%
	brush	0	0	159	18	216	1371	4268	299	408	431	0	7170	59.53%
	pine	0	0	27	2	1245	49	397	51071	3120	3283	1475	60669	84.18%
	decid	0	0	272	131	225	73	1296	4662	45252	5499	2101	59511	76.04%
	water	0	0	0	18	89	0	421	223	31084	9667	41502	74.90%	
	unclassified	0	0	0	0	94	0	0	1079	1935	395	4594	8097	56.74%
Total		0	0	954	3835	37185	12309	10126	62109	52717	41897	24772	245904	73.87%

(c) Using rule-based classifier with 7 bands with prediction importance over 2000

Classification Agreement

		Assigned											Total	% Agree
		burned	paved road	building	ground	crop	grass	brush	pine	decid	water	unclassified		
True	burned	133	22	0	0	1	0	14	56	113	11	106	456	29.17%
	paved road	7	110	54	11	18	91	161	9	73	0	2	536	20.52%
	building	0	10	227	29	0	103	26	0	32	0	0	427	53.16%
	ground	0	93	80	1805	0	134	209	13	108	0	0	2442	73.91%
	crop	1189	1034	48	8660	16335	9202	5264	1142	260	2970	3661	49765	32.82%
	grass	32	207	41	3287	2635	5337	3536	180	30	19	25	15329	34.82%
	brush	35	532	26	492	1797	1201	1804	364	897	15	7	7170	25.16%
	pine	12156	2461	15	0	1189	0	4098	17975	6087	1278	15410	60669	29.63%
	decid	772	8490	805	0	21	2	2586	8482	32570	1	160	53889	60.44%
	water	520	79	4	0	913	7	449	384	102	36690	2354	41502	88.41%
	unclassified	964	213	1	0	507	1	203	580	268	2377	2983	8097	36.84%
Total		15808	13251	1301	14284	23416	16078	18350	29185	40540	43361	24708	240282	48.26%

(d) Using minimum distance classifier with three Aerial and four Gabor bands

Figure 8: Confusion matrices using different classifiers and data layers.

distance classifiers to produce noisy results but the decision tree classifier was the most successful in overall classification.

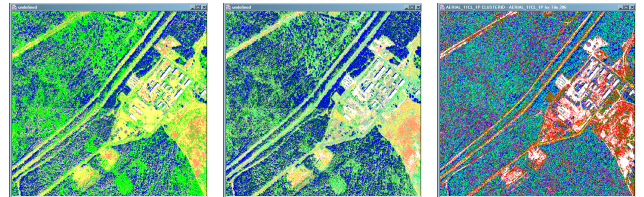
5.1.4 Evaluation of robustness to missing data

Figures 10 and 11 illustrate classification in the presence of missing data. Aerial and Ikonos bands were used in these examples. Missing data in the Ikonos bands initially resulted in many false alarms. However, land cover classification drastically improved when surrogate splits were used.

5.2 Region Level Classification

The initial evaluation was done using a synthetic image that contained regions with circular, elliptical, square and rectangular shapes. Grouping these regions into different sub-classes showed that ECCENTRICITY is useful for distinguishing compact shapes (square has a small value) from narrower and longer shapes (ellipse has a larger value and long rectangle has the largest value), THETA is useful for capturing rotation (vertical ellipse vs. rotated ellipse), INVARIANT_1 is useful for capturing circularity (square vs. circle), and SIGMA_X is useful for capturing vertical vs. horizontal elongation (rectangles).

Another scenario where region features are useful is the learning of roads. It is very hard to distinguish paved roads from parking lots using pixel level classification because of

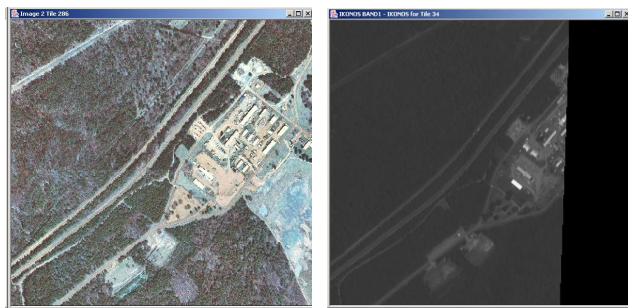


(a) Using decision tree classifier

(b) Using minimum distance classifier

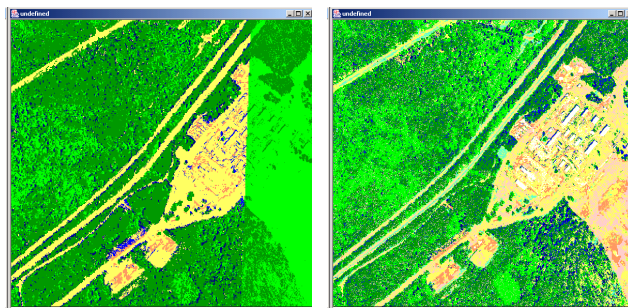
(c) Using k -means clustering

Figure 9: Classification using the decision tree classifier versus the supervised minimum distance classifier and the unsupervised k -means clustering using the three Aerial bands and the same training data. Most of the labels were not separated successfully using unsupervised clustering. Absence of Gabor features also caused most of the noisy results.



(a) Aerial bands

(b) Ikonos band with missing data



(c) Classification using no surrogate splits

(d) Classification using 5 surrogate splits

Figure 10: Classification in the presence of missing data. Missing data in the Ikonos bands (see Figures 2(c) and 2(d)) causes the classification to give false results for the right half of the scene. However, recognition of buildings as well as land cover classification drastically improves when surrogate splits are used.

their similar surface characteristics but roads have significantly different characteristics than parking lots in the region level. We trained a decision tree classifier for the classes “roads” and “non-roads” using the segmented Fort data (some of the example regions are shown in Figure 12). The resulting rule was:

```
IF SIGMA_MINOR < 2.488
  AND LABELID in {2}
THEN CLASS road WITH PROB 0.793701
```

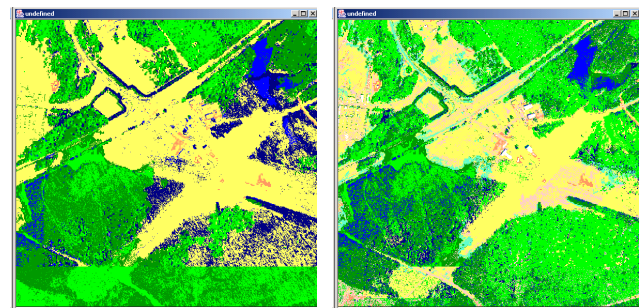
This rule shows that SIGMA_MINOR, i.e. the spatial variance of a region along its minor axis, distinguishes very narrow road regions from more rectangular parking lots. Note that the second condition in the rule requires that a road region consists of pixels classified as “paved” using the pixel level classifiers (see Table 1).

We also experimented with the statistical summaries of the pixel contents of regions in building region level land cover models. We computed the mean (ELEVATION_MEAN) and standard deviation (ELEVATION_STDDEV) of elevation for each segmented region using the elevation attributes of pixels inside that region. These two values are included



(a) Aerial bands

(b) Ikonos band with missing data and cloud



(c) Classification using no surrogate splits

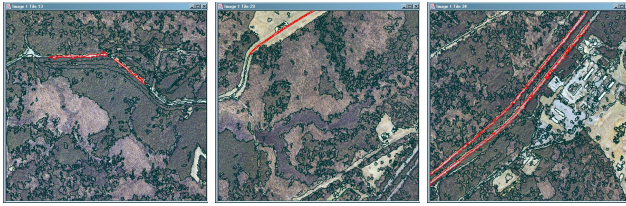
(d) Classification using 5 surrogate splits

Figure 11: Classification in the presence of missing data (cont.). This scene has missing data at the bottom and left borders within the Ikonos bands as well as clouds within the Ikonos imagery. When no surrogate splits are used, land cover classification is rough and inaccurate, also the missing Ikonos layers cause misclassification in the lower bottom of the scene. When surrogate splits are used, strong cloud shadows still lead to slight misclassification of forest in the left half of the scene but overall accuracy is improved.

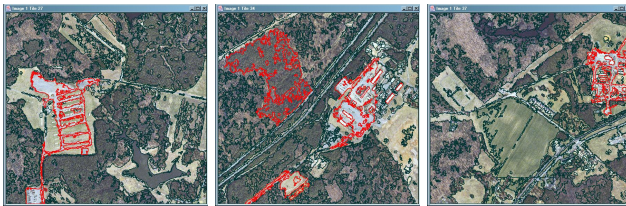
as new region level features in addition to the region shape features. We trained a decision tree classifier for the classes “flat fields” and “others” in a scenario of finding large and flat areas that are suitable for troop deployment. First, we chose large regions whose pixels were assigned to the “crop” or “grass” classes using the pixel level classifiers. Then, we input the regions that do not have hills in their neighborhood as examples for the “flat fields” class and the regions that were close to hills or lakes as examples for the “others” class. Some of the example regions and their corresponding elevation data are shown in Figure 13. The resulting rule was:

```
IF AREA >= 11452.5
  AND ELEVATION_STDDEV < 2.51875
THEN CLASS flat field WITH PROB 0.820335
```

This rule shows that the decision tree classifier automatically captured the importance of area and uniformness of



(a) Roads



(b) Non-roads

Figure 12: Learning roads using region-based training. Segmented region boundaries are marked with black. Example regions used in training are marked with red.

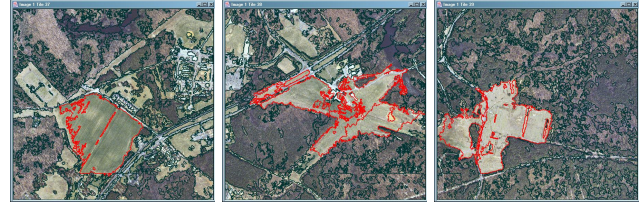
elevation (small standard deviation) from training examples. We believe that incorporating these region level rules with spatial relationships [2] (e.g. flat fields that are near a wide road that is connected to a village of interest) will provide remote sensing analysts with very valuable tools.

6. CONCLUSIONS

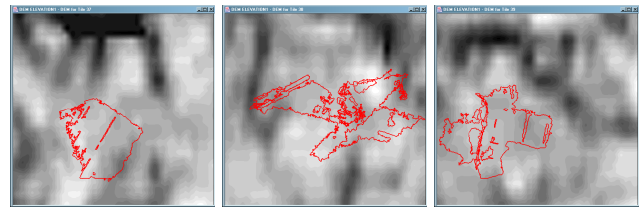
We described decision tree and rule-based tools for building statistical land cover models using our VISIMINE system for interactive classification and mining of remote sensing image archives. The decision tree classifiers and the corresponding rules proved to be very powerful in building land cover models by fusing data from different numerical and categorical data layers. They were also very robust to missing data and provided efficient land cover models with a small amount of training examples.

We compared the decision tree classifier with the rule-based classifier. Classification accuracy for generalized rules was very similar to (even slightly higher than) the one for the decision tree classifier. We also compared these two classifiers with the unsupervised k -means clustering algorithm and the supervised minimum distance classifier. The decision tree classifier appeared superior, not only in its capability to handle numerical as well as categorical data and to cope with missing data values, but also in regard to handling pure numerical feature data. The Gabor features that we used as an additional ancillary data layer were also quite useful especially in smoothing noisy Aerial data. Micro-texture analysis algorithms like Gabor features are useful for modeling neighborhoods of pixels and distinguishing areas that may have similar spectral responses but have different spatial structures.

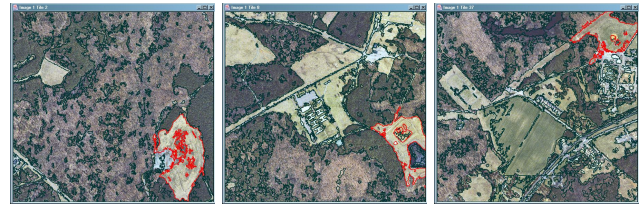
Our current work includes modeling interactions between multiple regions using region spatial relationships and using



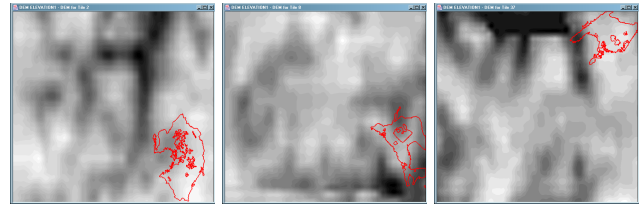
(a) Flat fields: Aerial data



(b) Flat fields: DEM data



(c) Non-flat fields: Aerial data



(d) Non-flat fields: DEM data

Figure 13: Learning flat fields using region-based training. Examples regions used in training are marked with red.

these relationships for mining image archives to find characteristic features that can distinguish different scene classes, and learning of decision forests as ensembles that use multiple decision tree classifiers. We believe that incorporating spatial relationships information into the region level rules will help modeling of higher-level structures that cannot be modeled by individual pixels or regions and will provide remote sensing analysts with very valuable tools.

7. ACKNOWLEDGMENTS

This work was done when Selim Aksoy was at Insightful Corporation and was supported by the U.S. Army contracts DACA42-03-C-16 and W9132V-04-C-0001. VISIMINE project was also supported by NASA contracts NAS5-98053 and NAS5-01123.

8. REFERENCES

- [1] S. Aksoy, K. Koperski, C. Tusk, and G. M. J. C. Tilton. Learning Bayesian classifiers for a visual grammar. In *Proceedings of IEEE GRSS Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, pages 212–218, Washington, DC, October 2003.
- [2] S. Aksoy, C. Tusk, K. Koperski, and G. Marchisio. Scene modeling and image mining with a visual grammar. In C. H. Chen, editor, *Frontiers of Remote Sensing Information Processing*, pages 35–62. World Scientific, 2003.
- [3] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth and Brooks/Cole, 1984.
- [4] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., 2000.
- [5] R. C. Gonzales and R. E. Woods. *Digital Image Processing*. Addison-Wesley, 1993.
- [6] G. M. Haley and B. S. Manjunath. Rotation-invariant texture classification using a complete space-frequency model. *IEEE Transactions on Image Processing*, 8(2):255–269, February 1999.
- [7] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*. Addison-Wesley, 1992.
- [8] X. Huang and J. R. Jensen. 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, October 1997.
- [9] K. Koperski, G. Marchisio, S. Aksoy, and C. Tusk. VisiMine: Interactive mining in image databases. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 3, pages 1810–1812, Toronto, Canada, June 2002.
- [10] P. Langley and H. A. Simon. Applications of machine learning and rule induction. *Communications of the ACM*, 38:55–64, November 1995.
- [11] R. L. Lawrence and A. Wright. Rule-based classification systems using classification and regression tree (CART) analysis. *Photogrammetric Engineering & Remote Sensing*, 67(10):1137–1142, October 2001.
- [12] R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. John Wiley & Sons, Inc., 1987.
- [13] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, August 1996.
- [14] M. Musavi, P. Natarajan, S. Binello, and J. McNeely. Knowledge based extraction of ridge lines from digital terrain elevation data. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 5, pages 2492–2494, 1999.
- [15] NASA. TERRA: The EOS flagship. <http://terra.nasa.gov>.
- [16] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [17] J. Rogan, J. Miller, D. Stow, J. Franklin, L. Levien, and C. Fischer. Land-cover change monitoring with classification trees using Landsat TM and ancillary data. *Photogrammetric Engineering & Remote Sensing*, 69(7):793–804, July 2003.
- [18] L.-K. Soh and C. Tsatsoulis. Multisource data and knowledge fusion for intelligent SAR sea ice classification. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 1, pages 68–70, 1999.
- [19] T. M. Therneau and E. J. Atkinson. An introduction to recursive partitioning using the RPART routines. Technical Report 61, Department of Health Science Research, Mayo Clinic, Rochester, Minnesota, 1999.