2000

Hierarchical feature extraction: A stepwise approach to image classification

Mark Vincent Mangrich
The University of Montana

Hierarchical Feature Extraction

A Stepwise Approach to Image Classification

by

Mark Vincent Mangrich

BS, Forestry, Iowa State University

Presented in partial fulfillment of the requirements for the

Degree of Master of Science

The University of Montana

2000

Approved by:

Chairman, Board of Examiners

Dean of the Graduate School

1-13-01

Date

Mangrich, Mark V., M.S., January 2001 Computer Science

Hierarchical Feature Extraction, A Stepwise Approach to Image Classification

Director: David W. Opitz DWO

Scientists frequently use remotely sensed digital imagery as a tool to measure and quantify the effects of global change. Accurate interpretation of digital imagery can make the difference between a correct and incorrect decision regarding our environment. The current techniques for image analysis are: (1) too slow to effectively interpret the volume of existing imagery, and (2) generally unsuccessful at accurately identifying cartographic features at a level that meets the expectations of the earth science community. Simplifying the image analysis process is an emcrging and pressing need; more imagery is becoming available and the cost of gathering the data is decreasing in price. Recent improvements in image interpretation are utilizing inductive learning algorithms. These systems show promise since they can process imagery quickly. However objects in images are very complex. It is difficult for inductive learners to identify complex features in an image with one model. The results are often incorrect and cluttered. We present an "assisted feature extraction system" that extends the simple one-pass inductive learning approach. The system applies a hierarchy of inductive learning algorithms that assist an analyst in interactively removing classification errors through a "data-driven" process. The basic idea is to tackle the classification task iteratively, reducing the problem into sub-problems that are more specific. Successive steps in the hierarchy eliminate extraneous concepts and leave the learner to focus on learning in an isolated problem area. The results are exciting since they show the hierarchical approach is more accurate than baseline single-pass machine learning algorithms at image classification.

Table of Contents

List of Figures

1. INTRODUCTION

Earth's environment is constantly changing. To keep abreast of the changes, scientists frequently utilize information that is in some way referenced to a geographic location (called geo-spatial data). Making appropriate use of geo-spatial data often involves the use of a Geographic Information System (GIS). In the strictest sense, a GIS is a computer system capable of assembling, storing, and manipulating geographically referenced information. Most GIS experts would probably agree that a large percentage of the value in a GIS lies in the geographic accuracy of its data. Remotely sensed digital imagery provides huge quantities of data that are both geographically accurate and current. Unfortunately most GIS systems cannot effectively extract the desired cartographic features from the volume of available imagery. *We present an "assisted feature extraction system" based on a hierarchy of inductive learning algorithms that allows a GIS analyst to iteratively improve the quality of identifying cartographic features.*

Feature extraction is the process of identifying the cartographic objects contained in an image, and is often performed together with image classification. Image classifiers separate features into distinct categories, reducing the feature space and facilitating object recognition in imagery. This is an important step in image analysis, but is often very time-consuming. Because the current demand for interpreted spatial data is unmet, simplifying the feature extraction task is a central problem in GIS. Analysts need new and innovative systems to extract information from imagery rapidly and with a high degree of accuracy.

Until recently, accurate image classification was only possible through the tedious method of 'heads up' onscreen hand digitization. As you can imagine, manually digitizing mass amounts of data is unfeasible. Consequently, automated techniques have been created to make large jobs possible. Many of these automated classifiers rely heavily on statistical or probabilistic classification approaches, and often focus only on spectral reflectance identification. The problem with this approach is poor performance on features that are highly dissimilar or do not exhibit an obvious spectral reflectance. For example, urban areas are relatively easy to identify spectrally in a scene, but spectral signature alone can't distinguish concepts such as the shape of a building or location of exit ramps along the freeway. These types of problems require spatial context. In addition, specialized feature finding algorithms only work for explicit features, and fail with varied parameters such as image resolution, change in season/time-of-day, geographic location, etc.

Inductive learning algorithms prove to be a viable alternative to current automated image classification processes (Bain, 2000; Burl et al., 1998; Maloof et al., 1998). Inductive learning is the approach of creating a program that is capable of generalization and can automatically improve its ability over time. Inductive learners take labeled examples and develop a model that predicts an output on unseen instances. The process begins with an initial training period, where the learner generates a hypothesis. Then the learner classifies the unseen input pixels in the image. Through empirical (real-world) examples, the learner actually generalizes at a level that custom fits the data, making inductive learning classifiers extremely adaptable to many problem types. In addition, these systems have the ability to recognize spatial patterns. Spatial pattern recognition permits more information to be interpolated into the decision process. This allows us to more closely simulate the way a human understands an image via spatial context. A final advantage of these systems is their ability to process imagery quickly. The inductive learning process described in this paper is 200 times faster than hand classification and 50 times faster than the process of hand classification and commercial image processing techniques (Opitz, et al., 2000).

A noticeable tendency of inductive learners is to over predict the frequency of a feature in the image, sometimes called the "false positive effect" (Palhang, et al, 1997). We often find that the learner incorrectly includes in the classification superfluous features along with correctly identified features. The cause of this problem stems from the fact that the learner is attempting to make a prediction for the features in the scene with one model. It is a difficult task to identify all the nuances of a complex feature in a single classification. The learner tends to include excessive clutter in the results more often than missing an object. Most baseline inductive learning systems do not allow the user to remove error after the classification stage, and unlike the hierarchical approach, these systems simply make one attempt at classifying the image. The user is unable to return to the classification in order to reduce errors. Errors introduced by "single-pass" machine learning approaches are present in the final results. One way to address this problem is to apply a hierarchy of inductive learners. This way the user and the system work together to improve the classification of image features.

Hierarchical feature extraction can iteratively remove clutter in a stepwise fashion and iteratively retrieve missing objects. The process begins with an intial classification, and then we iteratively reduce one type of error to an acceptable level. At this point the system can use a new learning algorithm to retrieve the missed objects. We can quit and accept the classification at any time. The system assists the analyst in identifying the targets in a scene quickly in hierarchical passes. It simplifies the classification process because we first identify general patterns and later focus our attention on classifying more specific patterns. For instance, suppose we are looking for all occurrences of primary highways in an image. A single pass learner may not be capable of discerning the difference between a primary road and secondary road. On the other hand, a hierarchical learner may be able to identify primary roads after an initial pass where we find all the roads. Now we only look within the road results to find the primary roads.

In order to demonstrate the flexibility and accuracy of the hierarchical classification method, we studied three diverse classification tasks: (1) a forest fire burn site in eastern Montana, (2) vehicle identification in an urban landscape, and (3) building detection in a high-resolution image of the Presidio in San Francisco. Each of the three classification tasks uniquely demonstrates potential applications for the system. We gain a substantial improvement in accuracy over one pass machine learning classifiers. In addition, the system performs very well classifying wildfire burn areas in eastern Montana.

In order to show the efficiency of an "assisted" feature extraction system, based on a hierarchy of inductive learning algorithms, this thesis is organized into sections on the background of the image classification task, methodology of the task, and results of the same. Finally, Section 6 discusses conclusions and future work for the project.

2. BACKGROUND

Image interpretation comes naturally to humans. We look at an image and immediately recognize our uncle, even if we have not seen him in years. However when asked to process thousands of images, the task quickly becomes overwhelming. That is why computers are used as a tool to help us interpret images. This section introduces and discusses the importance of image classification in the GIS community.

2.1. Image Classification Task

Image classification developed from the need to remove noise in geo-spatial data, allowing an expert to better interpret images. It evolved into quantitative techniques for automating the identification of features in a scene. Even before the launch of *Landsat1* in 1972, there has been great interest in the extraction of land use and land cover information from digital satellite remotely sensed data using digital image processing techniques (Lillesand and Kiefer, 1999). The objective of image classifiers is to categorize image features into classes or themes. The classified images produce thematic maps of the scene, which simplify image interpretation. The thematic maps are fed into a GIS database to be used by environmental decision-makers and GIS analysts. Figure 1 illustrates an example of image themes that an analyst may use in classification. Image classifiers can also be used to extract specific objects from images. This type of feature extraction is called object or target recognition. Multi-spectral and hyper-spectral imagery contain bands of data recorded in explicit spectral zones. Pixels in these images have a numeric value or digital number (DN) for every band. Even the most simple color images displayed on a computer screen contain bands of data.

Class No. (Color in Map)	Feature Type
1 (black)	Clear water
2 (green)	Dense Forest with closed canopy
3 (yellow)	Shrubs, Less dense forest
4 (orange)	Grass
5 (cyan)	Bare soil, built-up areas
6 (blue)	Turbid water

Figure 1: Example of image classes or themes

For example, a pixel in a simple color image has three DNs, one for the red spectral zone, one for the blue spectral zone and one for the green spectral zone. Multi-spectral and hyper-spectral images have many more bands than simple color images. These bands describe the image in many levels or energy spectral zones. By looking at the data bands wholly we can 'see' information invisible to the human eye. The power in this representation is that if used correctly, the bands contain a wealth of information about the image, and help to improve classification accuracy. However, in remote sensing image analysis, the difficulty arises in the fact that a feature type is often a conglomerate of objects or materials. As a result, the real world task of image classification is difficult because: (1) two objects may look entirely different but comprise the same feature type, and (2) spectral information alone is not enough to extract all features.

For example, suppose we want to identify all objects representing buildings in an image. It turns out that building detection is an extremely tough problem where a number of approaches have been tried (Gruen, 1997). One reason why is because the concept of "building" is a broad one, possibly ranging from irregular shacks to complex residential or commercial buildings all in the same image. Real world concepts like these are difficult because they are highly *disjunctive*. Direct application of commonly used image classifiers is problematic when processing these features, because the classifier must be capable of saying, “a building is a skyscraper” *or* “a building is a shack.” Confusion between skyscrapers and shack is often the undesirable result. Figure 2 illustrates a disjunctive feature ‘B’ in a simple two-dimensional plane.

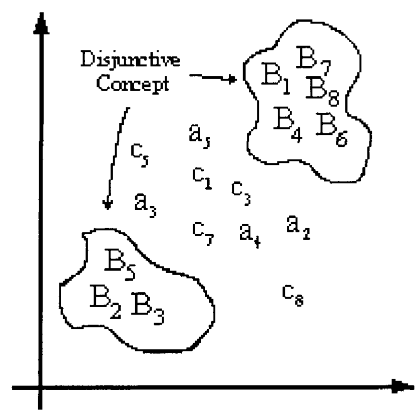
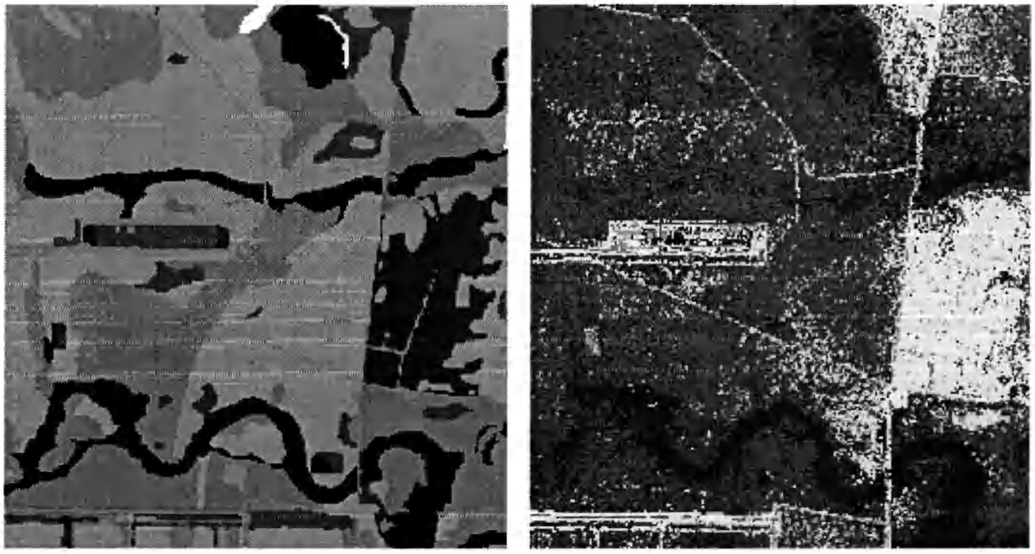


Figure 2: The disjunctive classification task

A second example illustrates the difficulty of feature extraction using only spectral information. Suppose our task is to discern the difference between asphalt roads, asphalt parking lots, and asphalt rooftops. The spectral signature of asphalt is very easy to identify, but spectral information alone makes it very challenging to separate these three similar concrete structures. In this case, space and context must be incorporated into the feature extraction process. The classifier must not look only at the spectral value for a pixel but also look at surrounding pixels to infer the notion of spatial perspective. Figure 3 gives an example to illustrate the importance of incorporating spatial context in the classification process.



(a) (b)

Figure 3: Two classification techniques for the same scene. Figure (a) uses spatial context, while Figure (b) does not.

Extracting complex multi-faceted features from an image made up of discrete pixels is a very challenging task. It is also a problem of great economic importance, because many real world decisions affecting each one us are made based on this information. Depending on the resolution, a pixel in a remotely sensed image can also be an aggregate of different material substances present on the ground. Despite considerable research effort, image classification remains a difficult task. As the information content of imagery increases, we need quality tools to autonomously determine and classify complex aggregate features.

2.2. Current Techniques

The suite of image classification tools available to the GIS analyst is impressive. Such products include automated or semi-automated techniques like "unsupervised",

"supervised", or "hybrid" approaches, and manual techniques such as the commonly used "heads up" hand digitization approach. Another option used to a lesser extent is to employ a computer specialist to write custom feature extracting software. These specialized algorithms can work great in context, but often perform poorly if the feature's characteristics change. For instance, a road finding algorithm that works well with images taken in the spring may need to be radically modified to find roads in the same area in the winter. Many analysts are finding that custom written classifiers are not versatile and too time consuming to write. In this section we will discuss a few of the most commonly used techniques and their effectiveness.

2.2.1. Hand Classification

Hand digitization, or hand classification, is probably one of the most universal approaches to feature extraction, especially to object recognition. It is often the most accurate approach available to many GIS analysts. In this process, a trained specialist manually traces the outline of a feature by clicking points on the screen. The difficulty is this process requires a great deal of time and skill to visually interpret just a few digital images, even when they are already placed in an orthographic map projection. For many applications this laborious process is unfeasible for the following reasons:

- The process is simply too slow to meet the demand for processed imagery.
- There is a lack of trained analysts available to perform the task.
- The cost of paying an expert to do this work by hand is prohibitive.
- The resulting quality diminishes as the expert becomes tired.

These problems have caused general users of GIS information to dismiss the accurate interpretation of imagery (Jensen and Cowen, 1997). To make matters worse, new high-resolution remote sensor imagery will be inundating the GIS community in the next few years. It has become obvious that hand digitizing techniques are incapable of accurately processing the volume information. So quantitative techniques for automating the identification of features in a scene must be used.

2.2.2. Automated Image Classification

Fortunately, we now have tools that can automatically classify images. Generally speaking, automated feature extraction techniques begin by, (1) defining the categories of interest (the features), (2) characterizing the categories using statistical analysis or inductive learning, (3) comparing each image pixel to the categories, making an assignment, (4) and finally, output the identified features. In this section three commonly used automated techniques are discussed.

Unsupervised Classification

Unsupervised classifiers group pixels in an image into classes based on the natural clusters present in the image values without the use of training data. The clustering methods create statistically defined spectral classes, of which the user has no prior knowledge. After the clusters are created, each pixel is classified based on statistical similarity to the clusters (Kramber and Morse, 1994). Finally the analyst identifies and labels each spectral class using existing information from ground maps, aerial photos, ground visits, or knowledge of the study area.

Since unsupervised classifiers do not utilize training data, the classes that result are based solely on the natural groupings in the image data. This can be convenient if the classes are not initially known, or if the analyst has trouble defining the classes in the image. However, problems arise in the following applications:

- The spectral classes are not linearly separable.
- The feature of interest is highly disjunctive (i.e., comprised of pixels from different spectral classes).
- The feature lacks highly distinctive spectral characteristics.

An example that illustrates these problems is a situation in which the computer defined classes have little informational value, like sunlit versus shaded conifers, or conversely if the spectral classes correspond to both conifers and deciduous trees (Lillesand and Kiefer, 1999). The process of unsupervised classification is too inflexible to cope with extracting complicated features in today's images.

Supervised Classification

Supervised classification methods use predefined training examples to generalize on unseen data. The process is often more accurate than unsupervised classification. The analyst begins by specifying the classes to be identified based on prior knowledge about the area shown in the image. The idea is to define the specific features of interest in the scene. Unlike the unsupervised classification process, the analyst is given the freedom to choose what classes are important. After the classes have been identified the analyst selects some representative pixels for training the classifier from each class. For example, let's say we are looking for the feature "conifers", we would identify examples of conifers in the scene and provide these as training data. During classification, each image pixel is assigned to a class based on either its statistical similarity to the training data, or the output of the inductive learning model.

One drawback to using the statistical approach to determining the similarity of a pixel to the training data is that it sometimes must assume a priori distribution of training pixels

(Mather, 1999). Inductive learning models do not have this requirement. They can accept a wider range of types of input data. Supervised classification is not perfect, and to address some of these problems, various techniques for hybrid classification have been developed.

Hybrid Classification

The objective of hybrid classification is to improve the accuracy of purely supervised or unsupervised procedures. Hybrid classifiers leverage the good qualities from both unsupervised and supervised classification, and thereby improve performance. There are many types, but basically a hybrid technique uses either supervised, unsupervised or both. An example of a hybrid classification sequence might be:

1. Carry out unsupervised classification of an image
2. Merge the pixels into raster polygons.
3. Manually delineate "Areas of Interest" (AOIs)
4. Select training data within those AOIs.
5. Associate spectral class values with training data.
6. Apply supervised classification to raster polygons within AOIs only.

Note: The concept of AOIs is an important one that will be addressed in Section 3.

2.3. Machine Learning

Machine learning draws on concepts from many fields: statistics, artificial intelligence, biology, mathematics, and others (Mitchell, 1997). The number of algorithms that fall under the heading machine learning is astounding, as it is a highly researched field (Langley, 1996). This thesis focuses on a subsection of machine learning algorithms called inductive learners. They infer (approximate) an output from a set of labeled inputs, i.e., inductive learners take labeled training examples and generate a model that predicts an output from unseen instances. Given a sufficiently large set of training examples and a proper hypothesis space, an inductive learner can successfully generalize beyond the observed data. This is a fundamental theory of machine learning (Mitchell, 1997). A learner's hypothesis space is the set of all possible predictions that can be made by a learner, and is used to describe the capability of the learning model.

Practice is required to effectively set up the problem so that a computer learns a concept effectively. The following three points represent decisions that must be made when applying machine learning to a classification problem.

(1) Training examples must be chosen so that they are representative of the concept to be learned, and so that they characterize the problem to be solved. Keep the idea in mind that the learner is capable of drawing a general conclusion from concepts present in the training data. Be sure that the training examples represent the output that you are trying to achieve.

(2) An output function (or target) must be chosen carefully because it determines the kind of knowledge that will be learned and how well the learner will perform. For example, if the goal is to distinguish the difference between a bird and a giraffe, then it is not important to concern the learner with what type of bird and how many spots on the giraffe (Langley, 1996, pp.187).

(3) The learning algorithm must be chosen with an appropriate bias for the job. Learning algorithms differ in the way in which they represent the hypothesis space and similarly the way in which they can represent target concepts. Each algorithm has its own bias. Without bias, the learner cannot generalize to anything outside the training set. Bias in machine learning is a good thing.

Our hierarchical feature extraction approach utilizes two inductive learning algorithms: artificial neural networks, and k-nearest neighbor. The artificial neural network (ANN) is a powerful learning algorithm that learns by adjusting weights in a network of interconnected web of nodes. After training, the inputs are fed into one side of the network, and a classification is given as output on the other side. The type of ANN we use is back-propagation. ANN learning is excellent for image classification because it is robust to noisy training data, and it can classify very quickly. Back-propagation is capable of creating a disjunctive output function, which is ideal for feature extraction. Figure 4 shows a general example of a 3 layer neural network with three hidden units.

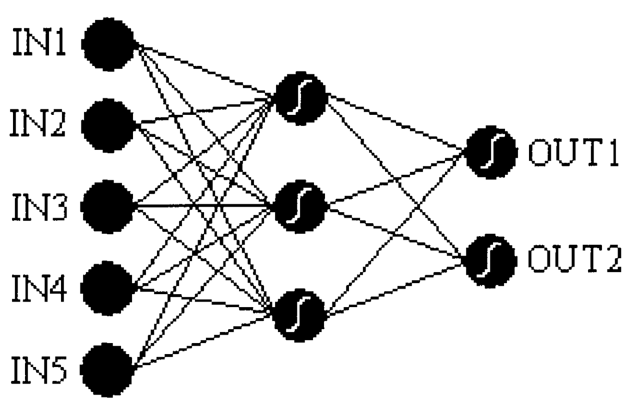


Figure 4: A 3 layer neural network with three hidden units.

K-nearest neighbor (KNN) is a widely used instance based learning method that works by classifying unseen examples based on the labels of the closest 'k' training instances (Mitchell, 1997). In contrast to ANN, which builds an output function during an initial training stage, instance-based learning algorithms create an output function dynamically for each new instance to be classified. This approach allows the learner to handle disjunctive concepts easily, and results show KNN can often out perform other learning algorithms at image classification (Maloof et al. 1998, Bain 2000). A drawback to this accuracy (there always is one) is that it takes a long time for KNN to classify. We will discuss how hierarchical learning can be used to reduce the classification time of KNN and increase the overall speed of image classification in Section 3. For more information on these algorithms please see Artificial Neural Networks, Chapter 4 of Mitchell's Machine Learning book, (Mitchell, 1997), and Instance Based Learning, Chapter 8 of the same book.

Problem Representation

How does one apply an inductive learner to the task of feature extraction? At its most basic level, we train a learner on a few example pixels, whose classification is known, and then classify every pixel in the image based on the prediction of the learner. This process is supervised classification. Pixels are inputs to the learning algorithm. Ancillary data such as elevation can also be inputs. The output or target is the classification. When classifying multiple band images, a pixel's input set to the learner is the set of spectral values. Each element in the set represents the pixel's DN for a band. If we have a 7-band image, in the simplistic approach, every pixel has an input set with 7 numeric values.

As we talked about in Section 2, many of the traditional image classifiers rely heavily on spectral information within the image. One of the problems with these techniques and with the simplistic representation listed in the previous paragraph is absence of spatial context when classifying. The representation used in the hierarchical technique incorporates spatial perspective. If we classify a scene by simply looking at a single pixel, we must base classification solely on *spectral pattern recognition*. In contrast, *spatial pattern recognition* incorporates into the classification geometric shape, and spatial context. Image classifiers that utilize spatial pattern recognition expand the set of input values to include surrounding pixels in a sliding window. Figure 5 illustrates an example of a sliding window; the center pixel's input set includes the surrounding 8 pixels for each band. This window is moved throughout the image, and is used to create the input set for every pixel.

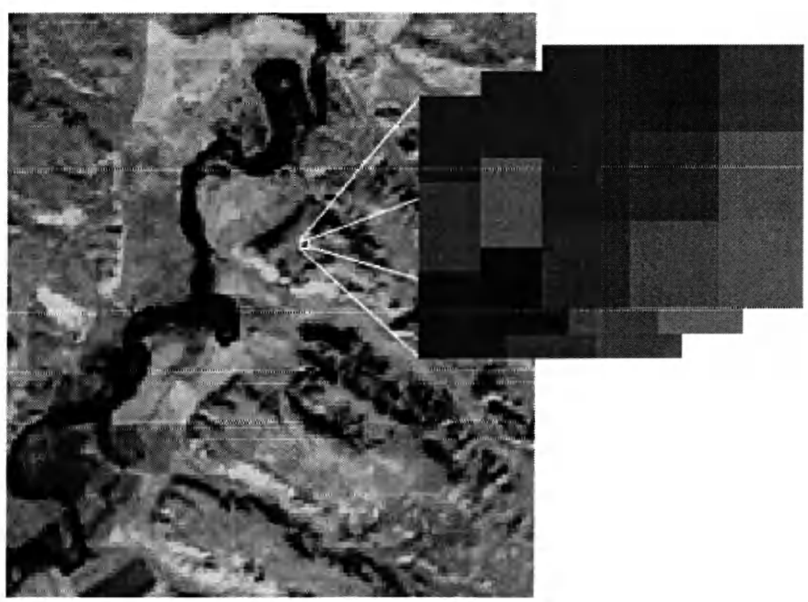


Figure 5: Example of pixel values in bands

All of the DNs in the input set describe the scene for that pixel spatially. This gives the reflectance value for the pixel spatial context to improve classification. The width of the sliding window can be fine-tuned to recognize shapes of different sizes. Unfortunately, as the size of the window gets large, the number of values in the input set can become overwhelming for the learner. In response to this problem, a convolution spatial filter that averages groups of pixels on the edges of the sliding window is often used to reduce the size of a pixel's input set (Bain, 2000). Modeled after human foveal vision, center pixels in this approach stay in focus while surrounding pixels are blurred in the periphery. Besides reducing the size of the input set, foveal representations may decrease noise present at maximum resolution, and may make the model inherently more scale invariant. In short, the foveal convolution filter is an example of any number of pre-processing image operations that can be performed to improve classification (Seul, et al. 2000). There are various ways to digitally manipulate images. Each technique pulls out certain image traits, and may improve the ability of a classifier to recognize a feature.

3. HIERARCHICAL APPROACH

With any classifier there is the potential for two types of error: (1) false positives and (2) false negatives. False positives, also called errors of comission or "clutter", occur when a classifier identifies an object incorrectly as the feature of interest. False negatives, or an error of omission, occur when an object is fully or partially missed by the classification. The goal of hierarchical feature extraction is to leverage a human's impressive vision ability to improve classification results by mitigating clutter (false positives), and retrieving false negatives. The overall process iteratively narrows the classification task into sub-problems that are more specific and well defined. Figure 6 illustrates the hierarchical approach.

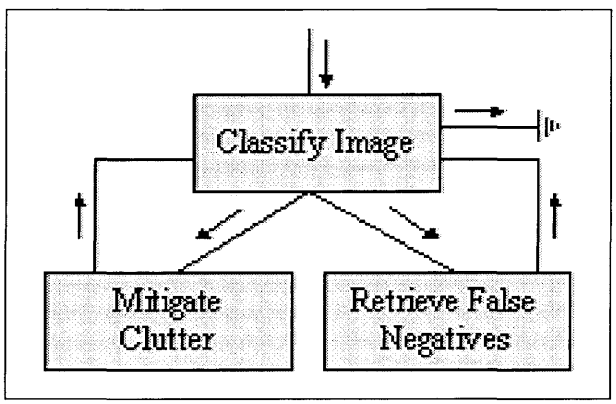


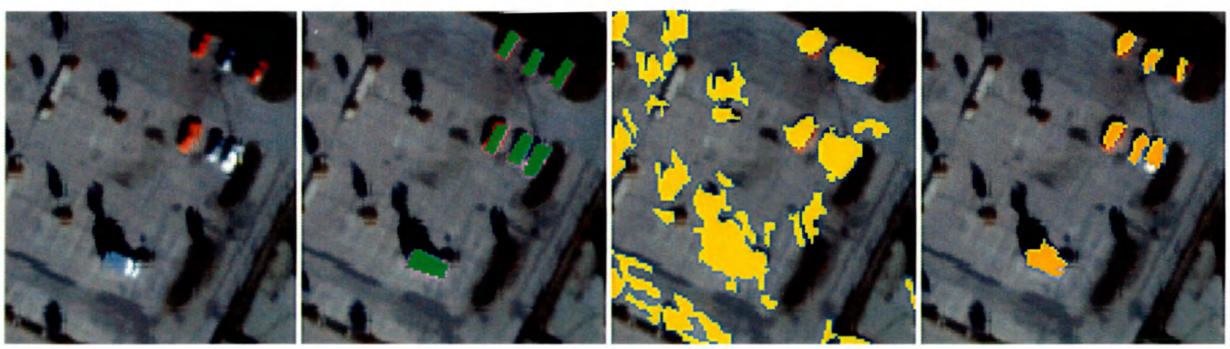
Figure 6: Hierarchical Feature Extraction Process

We begin the hierarchical process the same as we approach any baseline inductive learning classification, i.e., select labeled examples for the feature being extracted, train the learner, and then classify every pixel in the image based on the learner's prediction. At this point if we are not satisfied with the results, we can apply a hierarchy of learners to improve the classification. The lower two boxes in Figure 6 illustrate how we can iteratively reduce clutter and/or false negatives. The classification is improved in passes;

each new pass is designed to remove one form of error from the results of the previous pass. The concept goes back to the machine learning decisions in Section 2.3, i.e., correctly choosing training examples to emphasize a learning task.

Clutter Mitigation

Clutter is the most common form of error in automated feature extraction. The objective of clutter mitigation is to remove false positives, and thereby reducing the comission error. Thus, the learning task is to distinguish between false positives and correctly identified positives. The user generates a training set by labeling the positive features from the previous classification as either positive or negative. The trained learner then classifies only the positive instances from the previous pass. The negative instances from the previous pass are considered correct in clutter mitigation and are thus masked out. An example of clutter mitigation is shown in Figure 7. The classifier creates clutter by



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

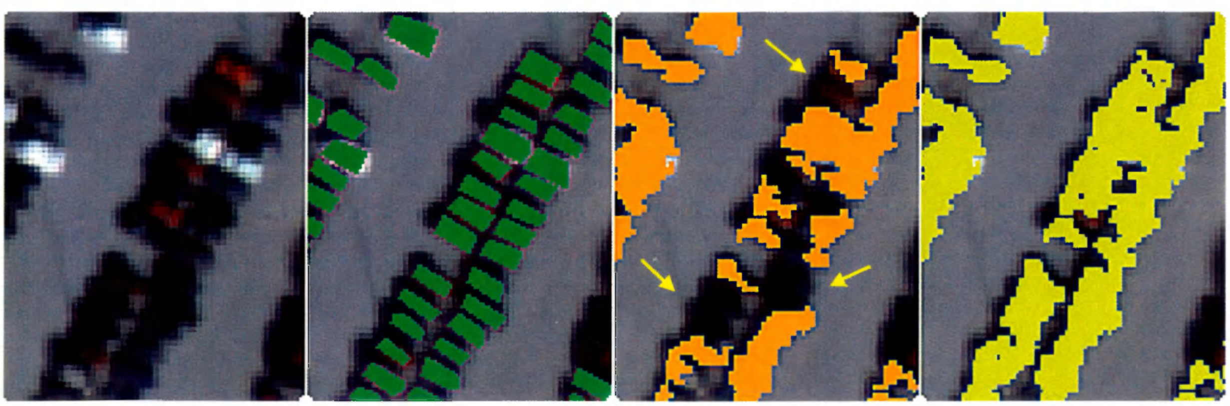
Figure 7: The classification on a section of an image: (a) Original unclassified image, (b) The true vehicle classification, (c) Clutter evident from Pass 1, (d) Clutter-mitigation results after Pass 2.

identifying small shadows as vehicles. The true classification of vehicles is given in Figure 7(b). Notice how the clutter in Figure 7(c) has been eliminated from the classification in Figure 7(d). We make a clutter mitigation pass between Figure 7(c) and Figure 7(d) to remove the small shadows from the classification.

False Negative Retrieval

The intention of false negative retrieval is to focus the learner on finding missed objects in the image, and thereby reducing the error of omission. Here, the learning task is to distinguish between false negatives and correctly identified negatives. The user generates a training set by simply selecting missed objects; the remaining negatives are assumed to be true negatives. The new learning task is more narrowly defined since the variability of the positives should be reduced; however, clutter could easily be reintroduced, which could later be mitigated.

Figure 8 illustrates the stepwise process of retrieving false negatives. Again, dark colored cars are very similar to small shadows. Notice in Figure 8(c) the dark vehicles are not classified. The training examples are selected from non-vehicle pixels within the image and from unclassified dark cars visible in Figure 8(c). We make a false negative retrieval pass between Figure 8(c) and Figure 8(d). The learner is trained to only classify the dark colored vehicles. The results from the object retrieval pass are combined with



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

Figure 8: Retrieving false negatives. (a) Original unclassified image, (b) true vehicle classification, (c) yellow arrows identify unclassified dark vehicles (d) final hierarchical classification.

the results from the previous pass via a logical 'or' function to give the results shown in Figure 8(d). The intent is to combine the results from both passes and pull dark vehicles into the classification.

Frequently in computer science, difficult problems are simplified to try to make some progress and gain insights, in hope that the insights can be then used to make progress on the original problem. As discussed in Section 2.3 above, choice of the target function characterizes the learner's ability to generalize. The idea of our method is to order the learning task so that we attempt large general problems first and then manage the more complex, specific concepts. One of the reasons why the process works is because features are inherently nested inside other features, e.g., dark cars, red cars, blue cars, and green cars are all cars. A specific feature is often a subset of a broad general feature, and we exploit this observable fact in the system to help the learner improve the results. Remotely sensed images often exhibit this general to specific ordering of features (Woodcock, et. al.1992). Thus, a common difficulty with single pass classifiers is their inability to completely identify complex features with a single model. The hierarchical technique addresses one of the major challenges to image classification, the problem of identifying aggregate disjunctive features, a problem we outlined in section 2.1. If a feature is composed of aggregate parts, possibly too disjunctive to classify with one model, then by means of this system, we mitigate clutter or retrieve false negatives to simplify the complexity of the feature.

The hierarchical process dynamically creates what the literature calls "Areas of Interest" (AOI) or a "Focus Of Attention" (FOA) (Burl et al, 1994). AOIs defined in the clutter mitigation pass can greatly reduce the number of pixels to be classified in the image. Recall that KNN dynamically generates an output function with each new pixel to be classified. Classification can be a slow process as KNN builds a new prediction for every pixel. The AOIs reduce KNN classification time. We can use the accurate KNN algorithm without the high costs associated with full image classification using the algorithm.

The process is "data driven", where as other statistically based supervised or hybrid classifiers often require an analyst's prior knowledge about a domain. The hierarchy is applied based on the constraints of the data. For example, suppose an expert analyst has learned to group hay and grass together as one feature because they are spectrally similar, and because he/she knows this will improve the results. Unfortunately, this technique requires the knowledge that hay and grass are spectrally similar. The nice thing about hierarchical feature extraction is that it will automatically find a logical alignment of features based on the data, eliminating the guesswork of ordering the feature extraction task.

4. METHODOLOGY

We choose three unique study areas to present our hierarchical classification system: (1) vehicle extraction in high-resolution imagery (2) building detection in an urban scene, and (3) classifying the severity of wildfire to vegetation. All three experiments demonstrate the systems ability to remove error and fine-tune the results. Moreover, the fire experiment is an excellent example of the system's ability to tackle a real-world problem defined externally by the US Forest Service. In this section we begin by describing general methods used for each experiment. Then in each specific section, we outline the nature of the classification task and how we approach solving the problem.

General Methods

This section is a general description of the parameters for all three experiments performed in this thesis. Training examples are selected by hand with as much accuracy as possible via visually interpreting the image. The outputs from the inductive classifiers are real valued numbers normalized between 0 and 255. The analyst chooses a threshold anywhere in this range to fine-tune the classification, (e.g., any output above 155 is positive and anything below 155 is negative). We use the standard values of learning rate of 0.1, momentum of 0.9, and a single hidden layer topology for ANN (Mitchell, 1997). The KNN is set to k=5 neighbors with inverse squared real-valued distance weighting. A rule based aggregation post processing step called "merge" is used to form groups of contiguous pixels with the same classification. We use the MEGA merge version with a similarity matrix of 0-255. The interested reader can consult Ford, 1997 for more information about "merge." The classification can be converted to an ordered set of eling tool, ESRI ArcView GIS. Due to the

specific methodology are separated in their

4.1. Vehicles Experiment

The vehicle experiment is conducted on images of the Presidio, a former military base in San Francisco. The scenes contain a near infrared (NIR) band along with red, green and blue (RGB) bands (shown in Figure 9 as grayscale). Here we are interested in demonstrating the ability of the system to improve on the single pass classification process.

Task

The classification task is simply to extract and classify all the vehicles in the image. As you can see from Figure 9, there are many forms of vehicles present in the scene. The concept "vehicle" is highly disjunctive. Vehicle recognition is a difficult task for two reasons: (1) all vehicles cannot be identified by 'one' unique spectral signature (e.g. red cars and blue cars are spectrally distinct), and (2) all vehicles do not have 'one' explicit shape (e.g., semi trailers are much larger than sport cars).

Specific Methodology

In the vehicles experiment, we follow the standard machine learning protocol of leaving a test set out of the learning process (Langley, 1996). Take two images 'A' and 'B', and make one of them the test image and the other the training image. Choose image 'A' to train and then classify on image 'B'. The classification on image 'A' is our training image results, and the classification on image 'B' is our test image results, which shows the learning algorithm's ability to generalize beyond the training set. We use a 5x5 sliding window input pixel representation. The training set is comprised of pixels that actually represent a vehicle and false positive pixels. The ANN classifiers run for 35 epochs.

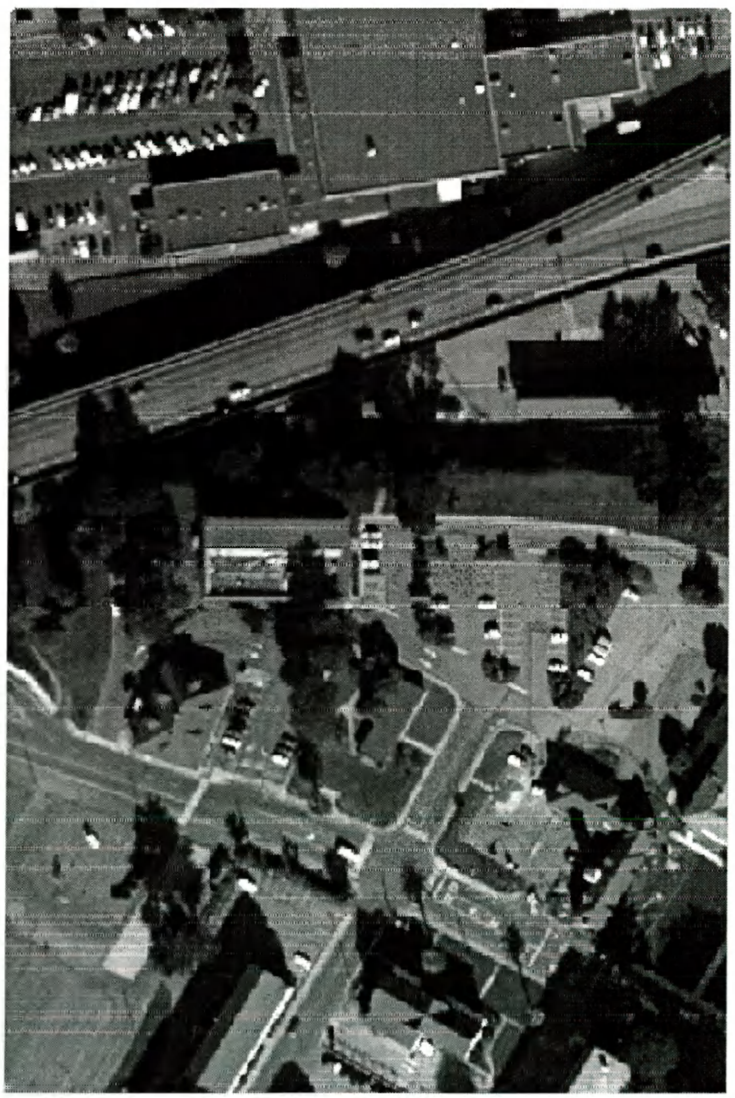


Figure 9: An urban section of the Presidio area used in the vehicles experiments

4.2. Building Detection Experiment

The building detection experiment was conducted on a residential area of the same Presidio imagery used in the vehicle experiments. Again the classifiers use the near infrared, red, green and blue bands contained in the image as inputs, with 8 bit pixel size. Figure 1 shows an area of the Presidio image used for this experiment.

Figure 1: A residential section of the Presidio area used in the building detection experiments.

Task

The classification task is simply to detect and classify all the buildings in the image. The image in Figure 1 is a good example of the complexity of the task. The buildings are oriented in many different directions and come in different sizes and shapes. One difficulty is the image contains concrete patios that look very similar to rooftops. This is a challenge for the learner.

Specific Methodology

In this experiment, we have two images; one a test set and the other a training set. The group of DN making up the inputs for a pixel is made up of a basic 3x3 sliding window, plus we use a foveal convolution filter to blur (average) eight 3x3 blocks surrounding the window. Counting the 4 input bands, this increases the cardinality of the input set from 36 to 68. Figure 11 illustrates the use of the foveal input representation. The idea is to size the sliding window so that it is proportional to the size of the buildings in the image. Similar to the vehicles experiment we use a high number of training examples, and in the case of ANN, a low number of epochs.

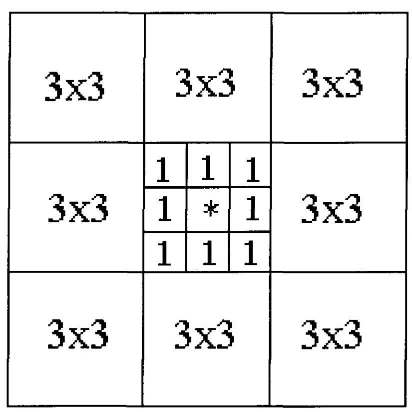


Figure 11: The input pattern for building detection experiment. Eight 3x3 blocks surround the center 3x3 window. The cells pixels in each of the eight surrounding 3x3 blocks are averaged to give one input per band.

4.3. *Fire Experiment*

This experiment presents a real world application of the system. Here we classify a Landsat-7 TM image of a wildfire burn site in eastern Montana taken on August 27, 2000 a few weeks after a 15,000-acre wild fire. The "fires of 2000" left many people wondering exactly how many acres burned and the severity of the damage to existing vegetation. The United States Forest Service requested a thematic map of the Fort Howes fire complex that burned near Ashland Montana in late July, 2000. The burned area actually spanned portions of two different Landsat-7 TM scenes, Path 35/Row 28 and Path 35/Row 29. Here we demonstrate the systems ability to find a difficult feature with out prior knowledge of how to classify a wildfire burn site. We are interested in using the hierarchical approach to improve the classification based on analyst interaction. The results were given to an image analyst for verification. He checked for errors of omission and commission using aerial photography. Any errors found were mitigated through the system.

Task

Our task is to thematically map the Landsat-7 image into the eight classes:

1. Unburned Tree
2. Mixed Burn Tree (mosaic burn)
3. Burned Tree
4. Burned Grassland
5. Unburned Grassland
6. Burned Shrub Land
7. Unburned Shrub Land
8. Barren

This study area is different from the previous two domains in that we had a sparse training set derived from air photo interpretation, and given to us by an expert analyst.

There were about 30 points provided for each class and from these few examples we classified the entire image. The concept "Mixed Burn Tree" is complicated since it is an aggregate of pixels from the classes "Unburned Tree" and "Burned Tree". The classifier must realize that Mixed Burn Tree is made up of some Burned Tree pixels and some Unburned Tree pixels, since wild fires can burn an area completely in one pixel and not burn the area in the next pixel. Making this task even harder is the fact that Burned Tree and Unburned Tree are in the classification task by themselves. When a class is comprised of other classes, some classifiers will label each pixel in mixed burn areas as either burned or unburned. The learner makes these types of classifications by using spatial context. Unburned Tree and Burned Tree are disjunctive concepts. Thus, the classifier must be capable of grouping these features into one feature. The concepts "Burned Grass" and "Burned Shrub" are also difficult to classify because both appear very similar at the 30-meter resolution of the Landsat-7 imagery. It is more difficult when the life forms are burned.

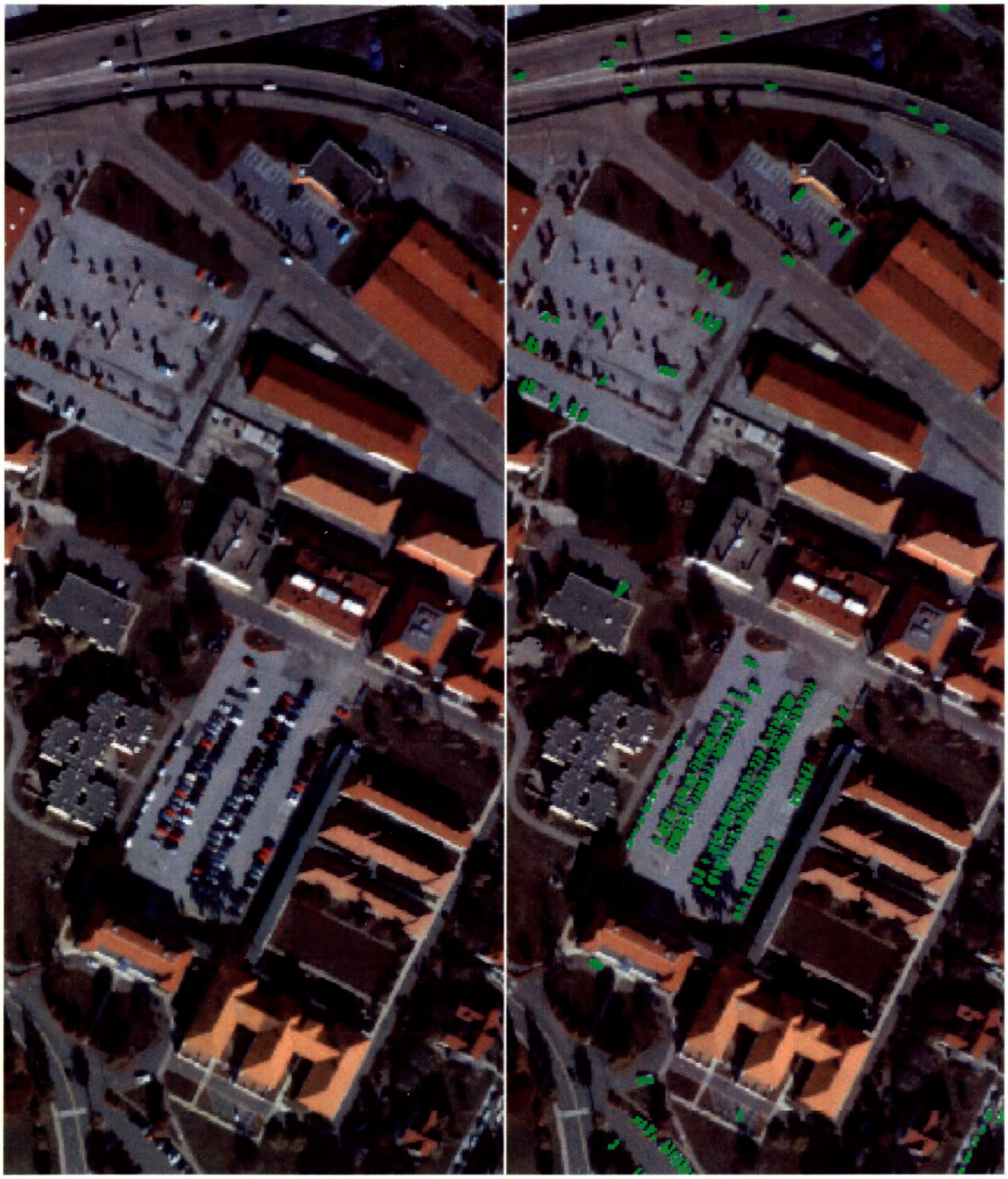
Specific Methodology

In this experiment, we do not follow the standard machine learning protocol of leaving an image out for testing since we only have one image, and few training examples. To increase the number of training pixels we expand each training point to cover 9 pixels (the center pixel plus 8 surrounding pixels). It is common for the features to be only 4 or 5 pixels across. Therefore, we use a sliding window input of 3x3. The idea is to size the sliding window small enough to take in account spatial context yet not too large to drown out these possibly small features. All eight classifications were extracted with a separate learner. The learner with the highest prediction decided the class for a pixel.

5. Results and Discussion

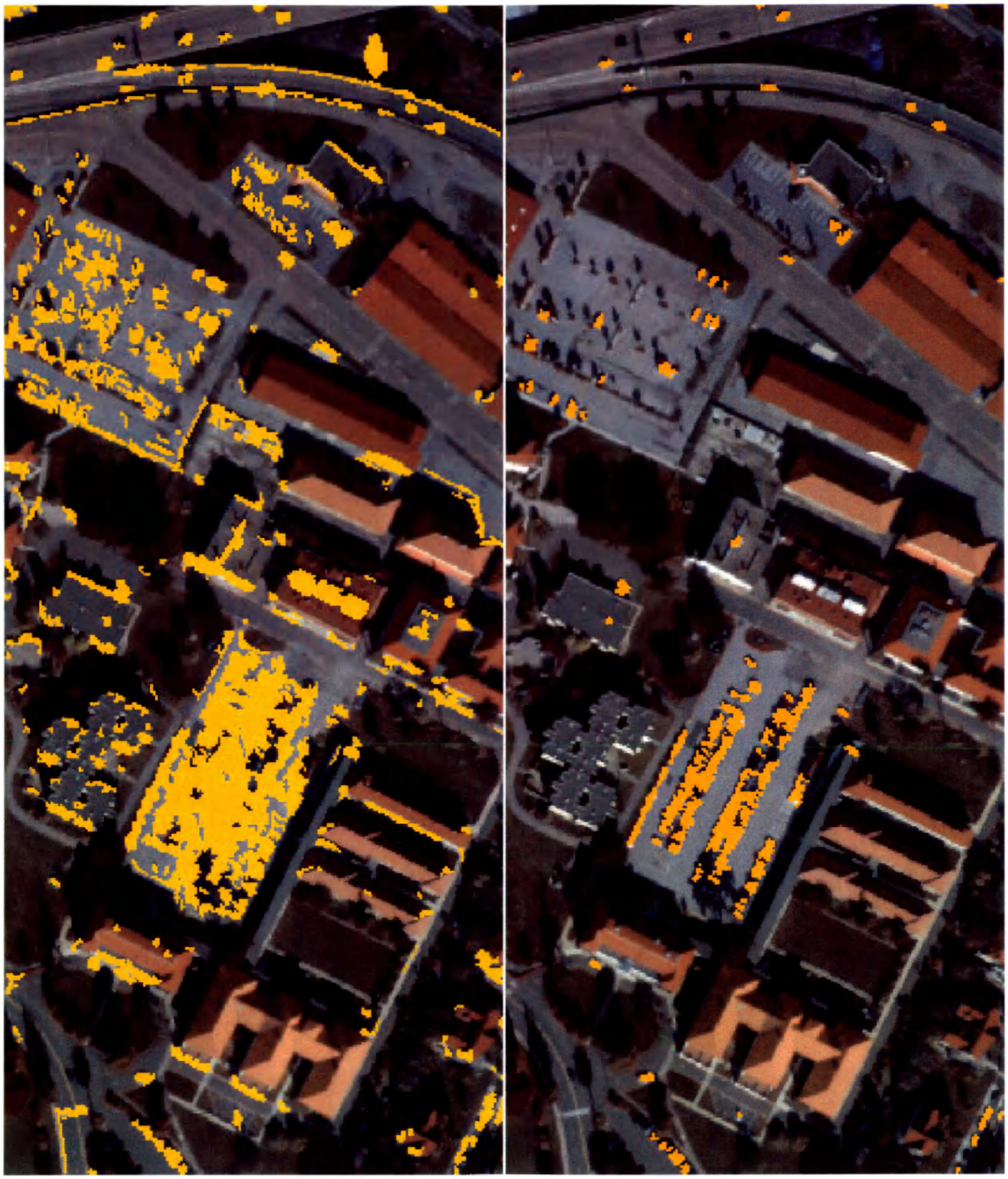
5.1. *Vehicles Experiment*

Figure 12 through Figure 14 show the training image results of the vehicle experiment. The images are laid out in consecutive order to illustrate how the hierarchical process progresses. The image in Figure 12(a) is an original unclassified section of the Presidio; each successive figure contains this same view. Something to note in this image is the similarity between small shadows and dark vehicles. The green classification in Figure 12(b) represents a carefully performed hand classification of the vehicles present in the scene. This classification was used for training. There may be error in this hand classification. Figure 13(a) starts the sequence of machine classifications. The results in this image are from a "single pass" inductive learning classifier. The high amount of clutter is apparent. Such clutter is typical when single pass techniques attempt a difficult concept such as vehicle detection. Following the hierarchical sequence we apply a clutter mitigation pass. Figure 13(b) illustrates nicely how one additional pass can successfully reduce a large quantity of false positives; however, note there are some vehicles (mostly dark colored ones) that are excluded from the classification during this pass. In the third step we want to retrieve the missed vehicles. Figure 14(a) contains the results after an attempt at retrieving missed objects. This step retrieves dark vehicles missed in previous passes. Note the vehicles on the highway at the top of the image and the vehicles in the parking lot at the center of the image. Also note this step reintroduces clutter that was mitigated in the previous stage. The solution applies a second clutter mitigation pass to the image, providing us with our final classification in Figure 14(b).



(a) (b)

Figure 12(a): An unclassified section of the Presidio Study Area. Figure 12(b): A carefully hand classified image delineating the vehicles in the scene.



(a) (b)

Figure 13(a): The training image of a 'Single Pass" ANN classification illustrates the clutter often present with single pass classifiers. Figure 13(b): The clutter mitigation results after pass 2 filtering. Notice how well errors of comission are reduced.

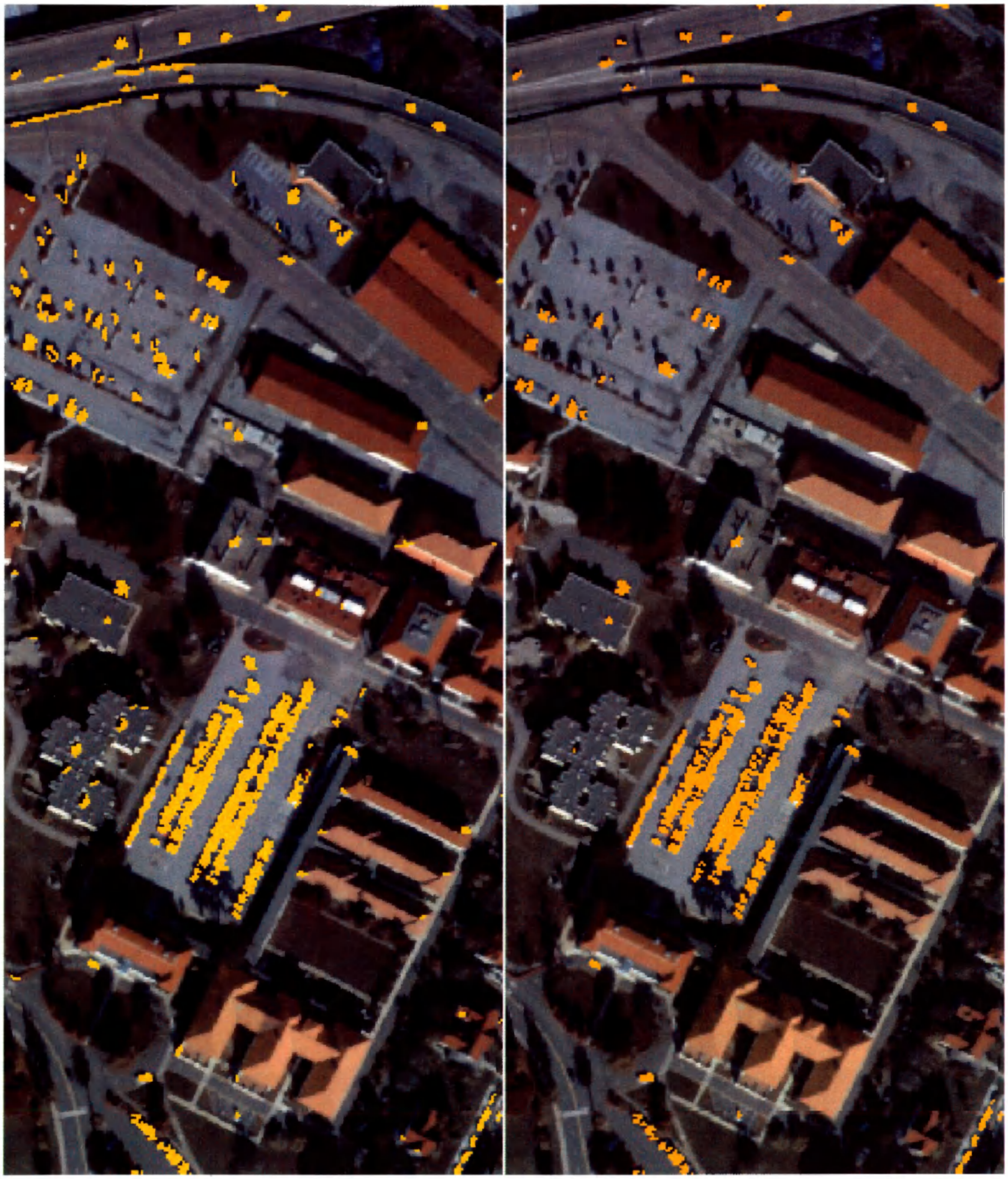


Figure 14(a): illustrates the results of Pass1 false negative retrieval performed after clutter mitigation. Figure 14(b): Final classification after four passes.

The number of true vehicle pixels in the Presidio Study Area (see Figure 12(b)) is relatively small compared to the number of non-vehicle pixels in the image. This can make accuracy reporting difficult since a completely unclassified image is 99% accurate. We present our results using Receiver Operating Characteristic (ROC) curves (Swets 1988). These curves offset the effects of data that is highly negative. The true positive rate is the ratio of correctly classified positive pixels to the number of total possible positive pixels. The false positive rate is the number of false positives (pixels that were classified as vehicles but really were not) over the number of pixels that are not vehicles in the image. The ratios may be easier to see in Figure 15. All pixels in the image that are truly vehicles are in the "yes" column under the truth heading. These "vehicle pixels" may be either correctly classified (true positives) or incorrectly classified (false negatives). All of the pixels in the "no" column under the truth heading are considered truly non-vehicle (all the pixels un-classified in Figure 12(b)).

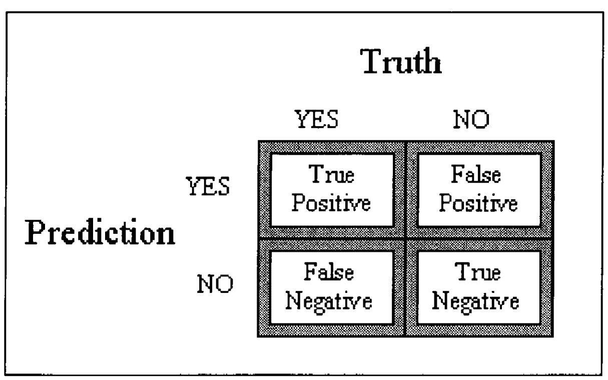


Figure 15: A confusion matrix

A ROC curve is often used in decision-making to compare the effects of adjusting a decision parameter. The outputs of a target function or decision model are plotted based on sliding this decision parameter. In this thesis ROC curves assess the effectiveness of each classification pass in the hierarchy. We adjust the threshold level on the real valued outputs of the learning model from 0 to 255. Figure 16 plots the test set results for the images in the vehicles study area. The same plot is shown in Figure 17. We scale the horizontal axis in Figure 17, to highlight the differences between each pass. The goal is to push the curve to the upper left corner of the plot area. That would indicate a low number of false positives and a high number of true positives. The ideal curve has a false positive rate of 0 and a true positive rate of 1. We select the best results from each pass as input to the following pass. For example in these results, ANNPass1+ANNPass2 is slightly more accurate than ANNPass1+KNNPass2. Therefore we show the next pass with ANNPass1+ ANNPass2. The peak in ANNPass1+ ANNPass2+ANNPass1 is evident in Figure 17. This plot shows the following succession: initial "single pass" classification (ANNPass1), clutter mitigation (ANNPass2), false negative retrieval (ANNPass1), and a final clutter mitigation pass (ANNPass2). The horizontal axis in Figure 17 is scaled to highlight the effects of applying classification passes. KNN on the final pass gives us the highest

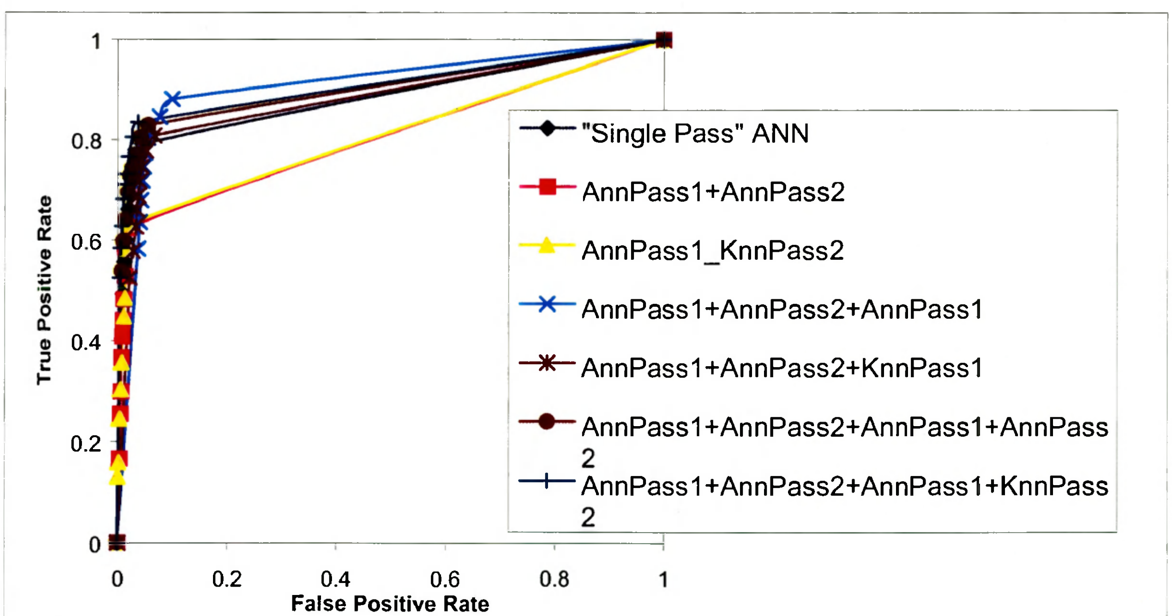


Figure 16: ROC comparison of vehicle classifiers. This chart shows the relationship between passes of the hierarchical learning system.

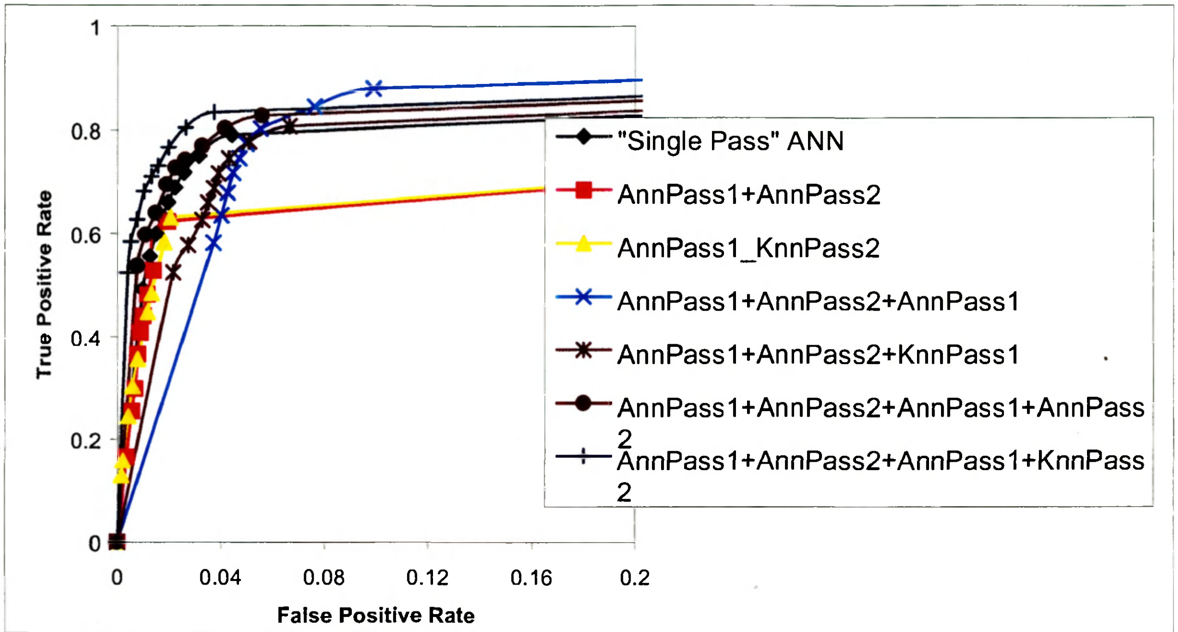


Figure 17: ROC curves of vehicle classifiers where the horizontal axis is scaled due to the high number of negative instances in the image.

true positive rate and the lowest false positive rate. Figure 18 compares the area under each ROC curve. Clutter mitigation (ANNPass2 or KNNPass2) passes tend to reduce the number of true positives while decreasing the false positive rate, hence the low values in the two initial clutter mitigation passes in Figure 18.

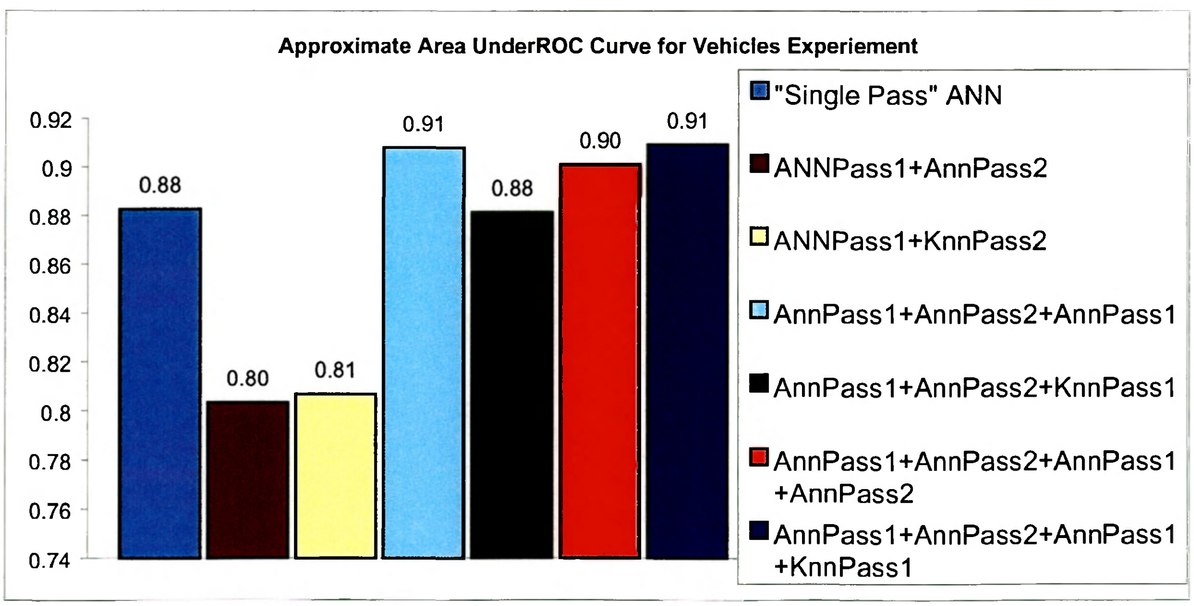


Figure 18: Approximate area under ROC curves for the vehicles experiments by "Pass"

5.2. Building Detection Experiment

In this experiment, we also compare the hierarchical classification process with a "single-pass" inductive learning classification of the same scene. Here, the single pass classifier has trouble discerning the difference between small concrete patios and rooftops. Unlike small patios, buildings in the image have a telltale ridgeline running down the center, and often have a shadow to one side. The clutter mitigation filter removes patios from the classification presumably by keying off these slight differences. Figure 19 through Figure 24 illustrate the training image results. Again, the images are laid out in consecutive order to illustrate the hierarchical sequence. The image in Figure 19 is an original unclassified section of the Presidio; each successive figure contains this same view.

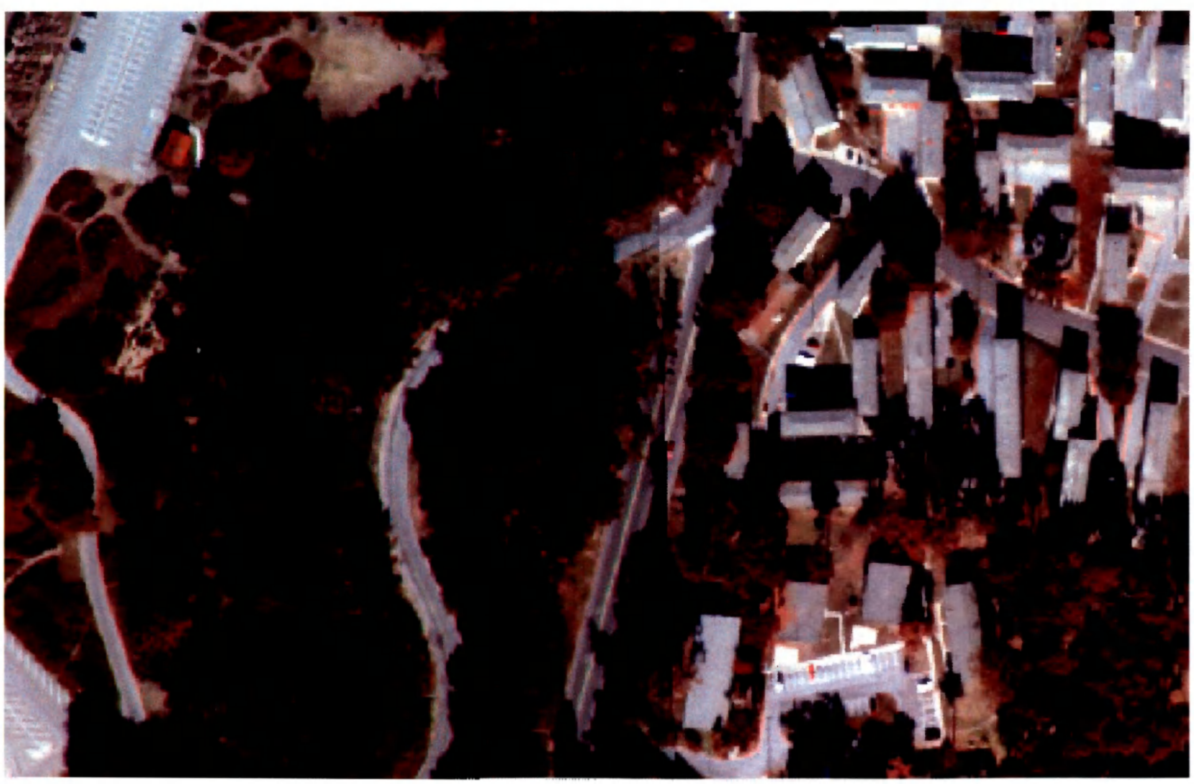


Figure 19: An unclassified section of the Presidio Study Area containing buildings.

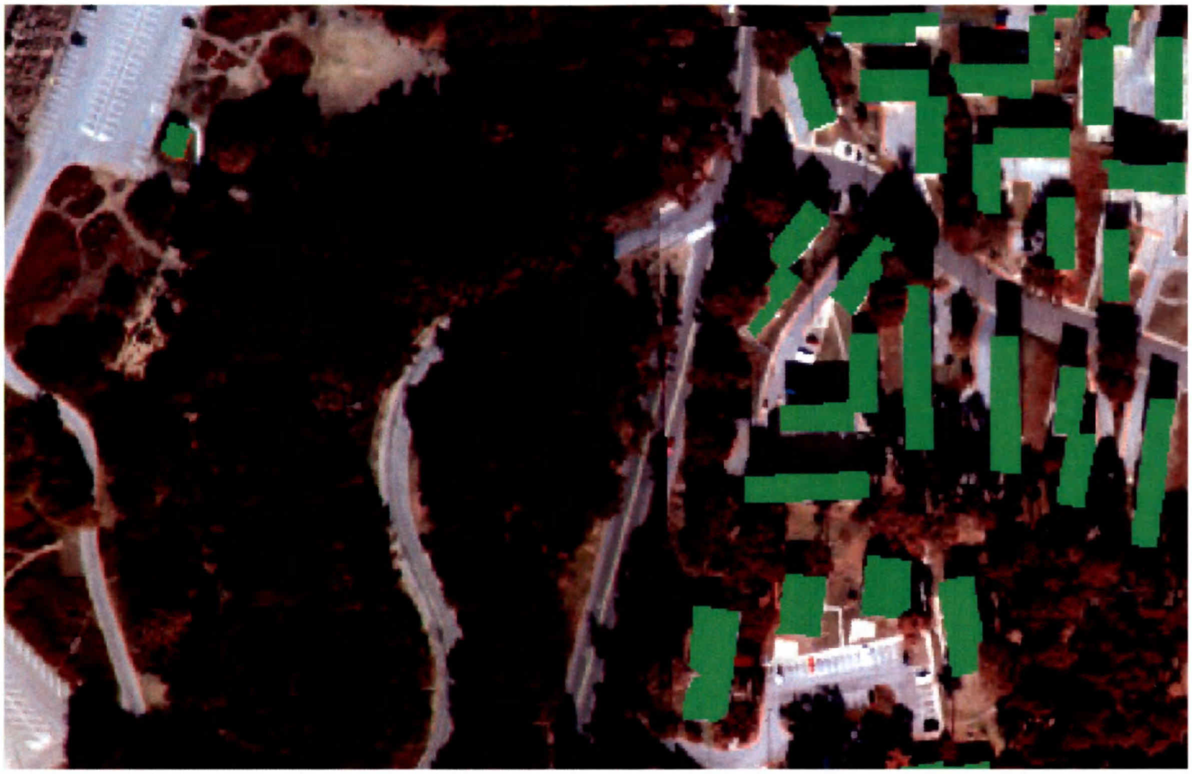


Figure 20: A careful hand classification of buildings present in the Presidio scene.

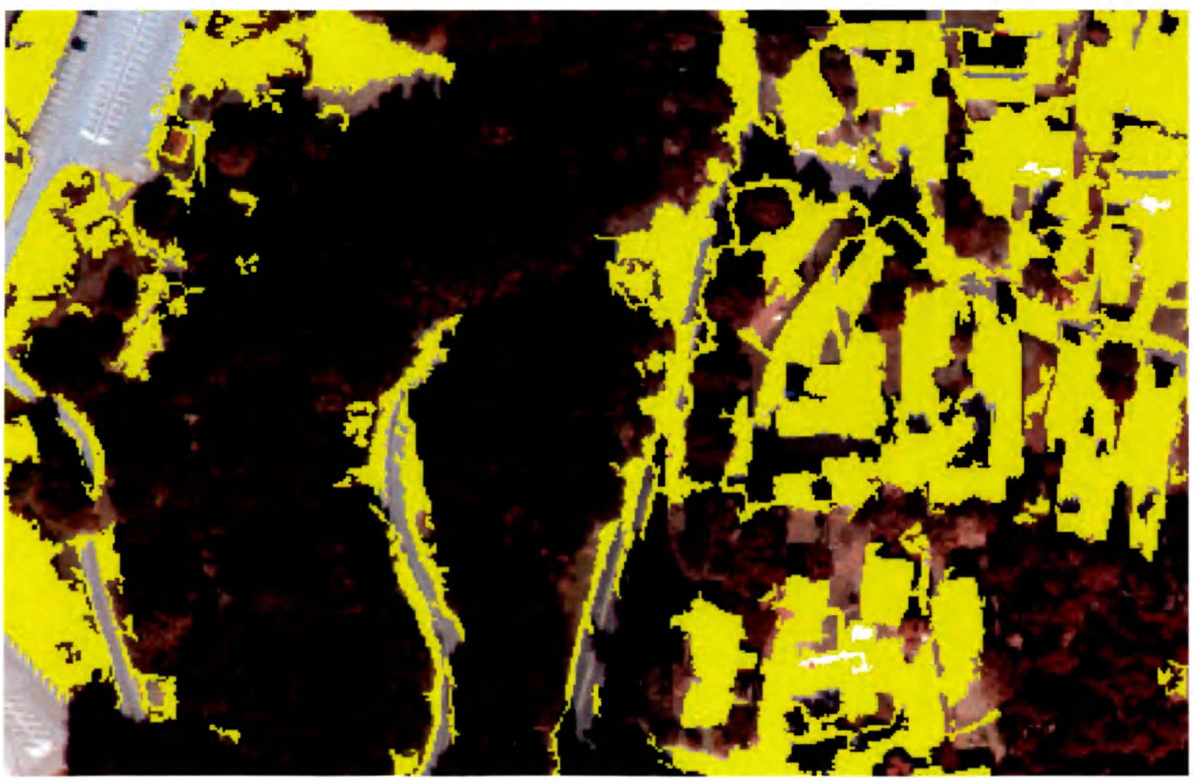


Figure 21: Building detection results with 'Single Pass' inductive learning classifier (ANN).

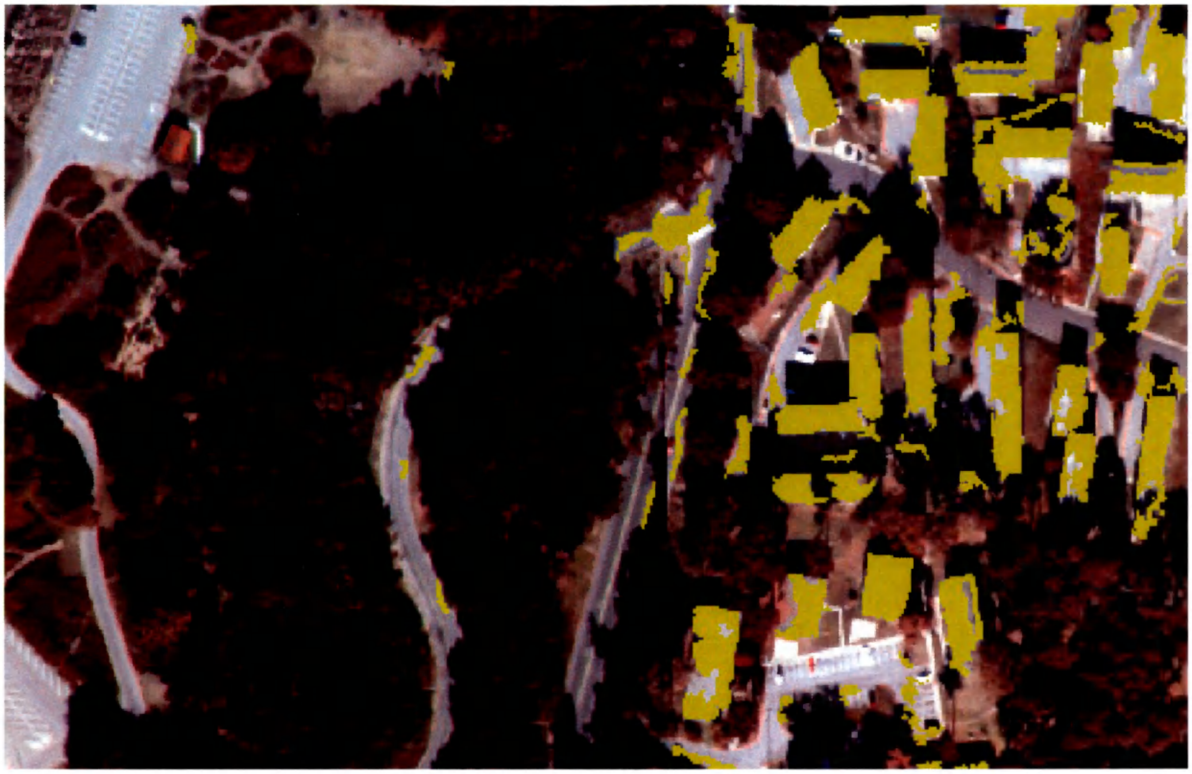


Figure 22: Initial clutter mitigation filter. Notice the mis-classified building upper left.

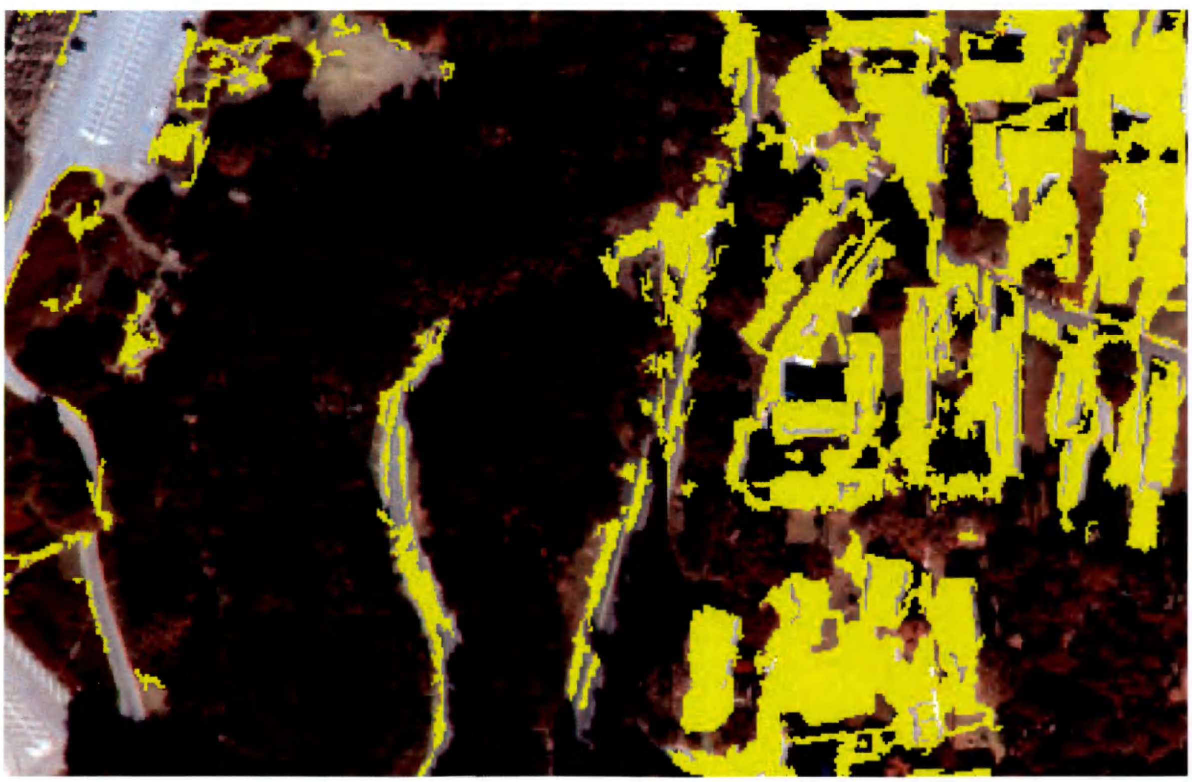


Figure 23: False negative retrieval pass. Notice the correctly classified building in upper left.

Figure 24: Final training image building detection classification. every building is identified

This particular sequence of images is interesting since the outlying building in the upper left hand corner of the scene is mis-classified in the initial pass. Subsequently, Figure 23 shows the correctly classified building after the object retrieval pass. The final classification in Figure 24 contains low clutter and every building is identifiable.

Again results are shown as ROC curves. Figure 25 relates the passes in the system. Our KNN final pass again provides the best classification (shown in yellow), and improves over the "Single Pass" ANN. The ROC curves in Figure 25 are shifted to the right slightly more than the ROC curves in Figure 15. This is due to the fact that in these experiments a greater percentage of the image pixels are true positives.

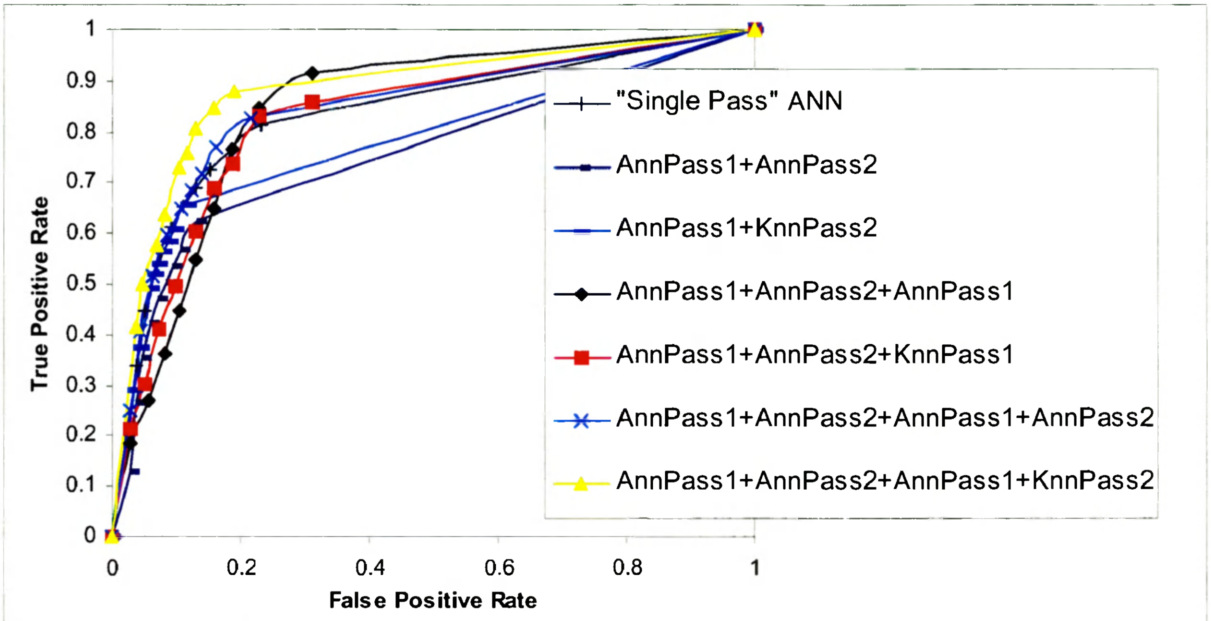


Figure 25: ROC curves of building classifiers. This chart shows the relationship between passes of the hierarchical learning system.

The ROC curves in Figure 25 reflect a higher number of true positive pixels than in the vehicles experiment. Figure 26 compares the area under each ROC curve. The final clutter mitigation pass has the highest area under the curve for all the classifications on the image. It is interesting to note that the initial clutter mitigation pass greatly reduces the area under the "Single Pass" ROC curve. The clutter mitigation pass greatly reduces the number of false positives in the classification. Since there are so many false positives in the image the effect on the curve is minimal. As noted above, clutter mitigation can also reduce the number of true positives. Since there are not as many positive pixels in the image as negative pixels, a slight loss of true positives is magnified. The effect is a noticeable change in the ROC curve.

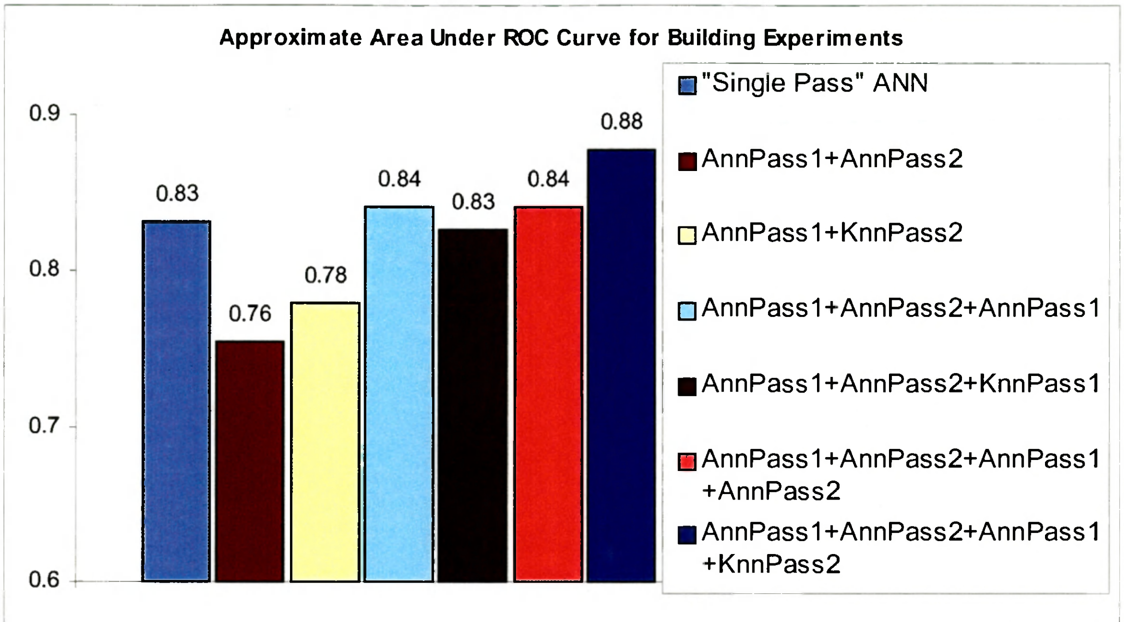


Figure 26: Approximate area under ROC curves for the building detection experiments by "Pass"

Classification Speed vs. Classification Accuracy

With all image classifiers there are two primary concerns regarding the effectiveness of the feature extraction system: (1) time the analyst needs to accomplish the classification, and (2) accuracy of the classification processes with respect to a careful hand classification or ground truth. World-class analysts at NASA's Jet Propulsion Laboratory (JPL) using a combination of commercial image processing techniques and hand classification also performed the building detection experiments given in this thesis. JPL's feature extraction technique is performed in three stages.

1. Register the Presidio imagery with IFSARE (Interferometric Synthetic Aperture Radar-Elevation) 2.5m-resolution images and DEM (Digital Elevation Model) 5-meter imagery to provide 3-D visualization of the image.
2. Combine the blue and near infrared bands to reduce shadow effects. Then apply a Photoshop "Poster Edges" filter to identify AOI.
3. Finally, manually delineate buildings by hand. (In lies the majority of analyst time.)

Since building tops look spectrally the same as streets and parking lots, JPL reported that the thematic classification presented difficult challenges. They noted that conventional multispectral classification techniques were therefore not very helpful in this respect. The classification performed by analysts at NASA JPL required an entire man month (160 hours) to complete. Analyst time in our approach is the time it took us to setup the classification process for the images, run two clutter mitigation passes, run a false negative retrieval pass, and to fine tune the system. Our approach takes about 200 minutes (3.3 hours) to classify the same image as JPL. The process described in this paper is 50 times faster than the NASA JPL classification (Opitz et. al., 2000). This is particularly exciting since the accuracy of the hierarchical process is comparable to the hand classification work performed by the experts at JPL.

5.3. Fire Experiment

An expert analyst provided the quantitative measure of accuracy used in this experiment. Figure 27 shows the unclassified Landsat-7 image used in the Ashland fire study area with bands 4, 5, 3 assigned to red green and blue.. The image is shown in false color. Orange is living vegetation, and ironically the darker the green the more stressed the vegetation. Figure 27 shows the results of the system to identify the perimeter of the fire. Each successive pass brought the classification closer to the expectations of the analyst. The fire perimeter underwent a clutter mitigation filter and a false negative retrieval pass. An inductive learner initially classified each of the seven features in the image. Input from the analyst was then used to reduce error evident in this initial pass through either a clutter mitigation filter or a false negative retrieval pass. Through the use of high-resolution aerial photography, taken of the area soon after the fire, we assessed the accuracy of subsequent passes. Our analyst believes the classifications for the eight features are very accurate. He did notice slight errors, however, in the "Burned Shrub", "Unburned Shrub", and "Barren" features, but these mistakes are easily cleaned-up by hand. The analyst is confident in the classification at a level consistent with the imagery available.

The concept "Mixed Burn" is a difficult feature to identify. We will use mixed burn to illustrate the hierarchical process for the other seven classes. Results from the initial pass erroneously identify senescent vegetation outside the burn perimeter as mixed burn. We use a clutter mitigation pass to remove these areas from the mixed burn classification. Figure 29 illustrates the effects of the clutter mitigation filter. The initial classification is shown in Figure 29(a).

The yellow circles in Figure 29(b) point out the effects of the clutter mitigation filter. A second clutter mitigation pass was not needed. The other seven features in the image were identified similarly. Only concepts unburned shrub, unburned tree, and unburned grass required a second and final clutter mitigation pass. Interestingly, these concepts are complex because they comprise the highly disjunctive group of unburned features in the image. Recall that the learner's prediction is in the form of real valued outputs in the range from 0-255, and that all eight classifications were extracted with a separate learner. The class with the maximum prediction decides the classification for each pixel.

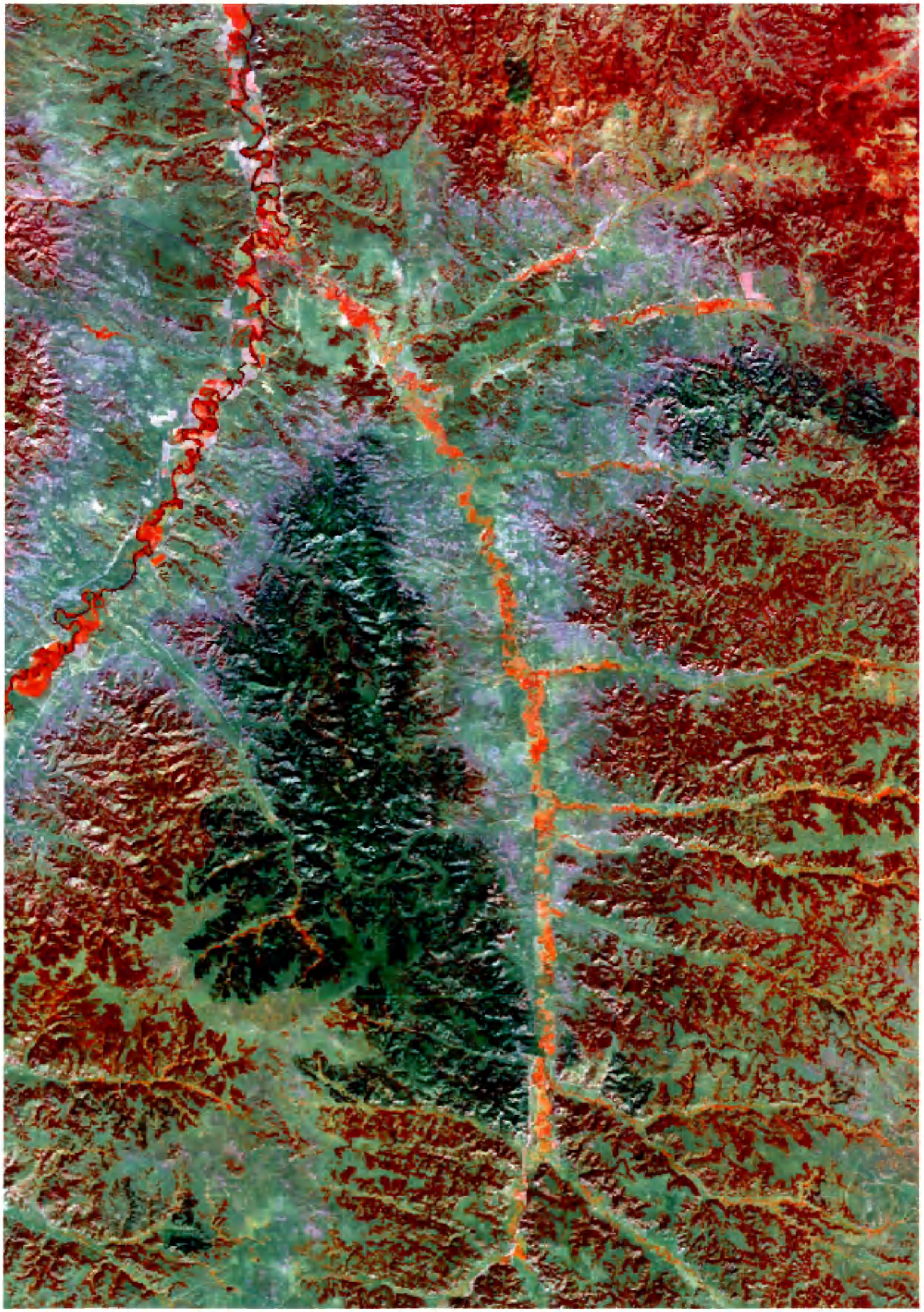


Figure 27: LandSat-7 TM image in false color, bands 5, 4, 3 (R, G, B) for fire experiments.

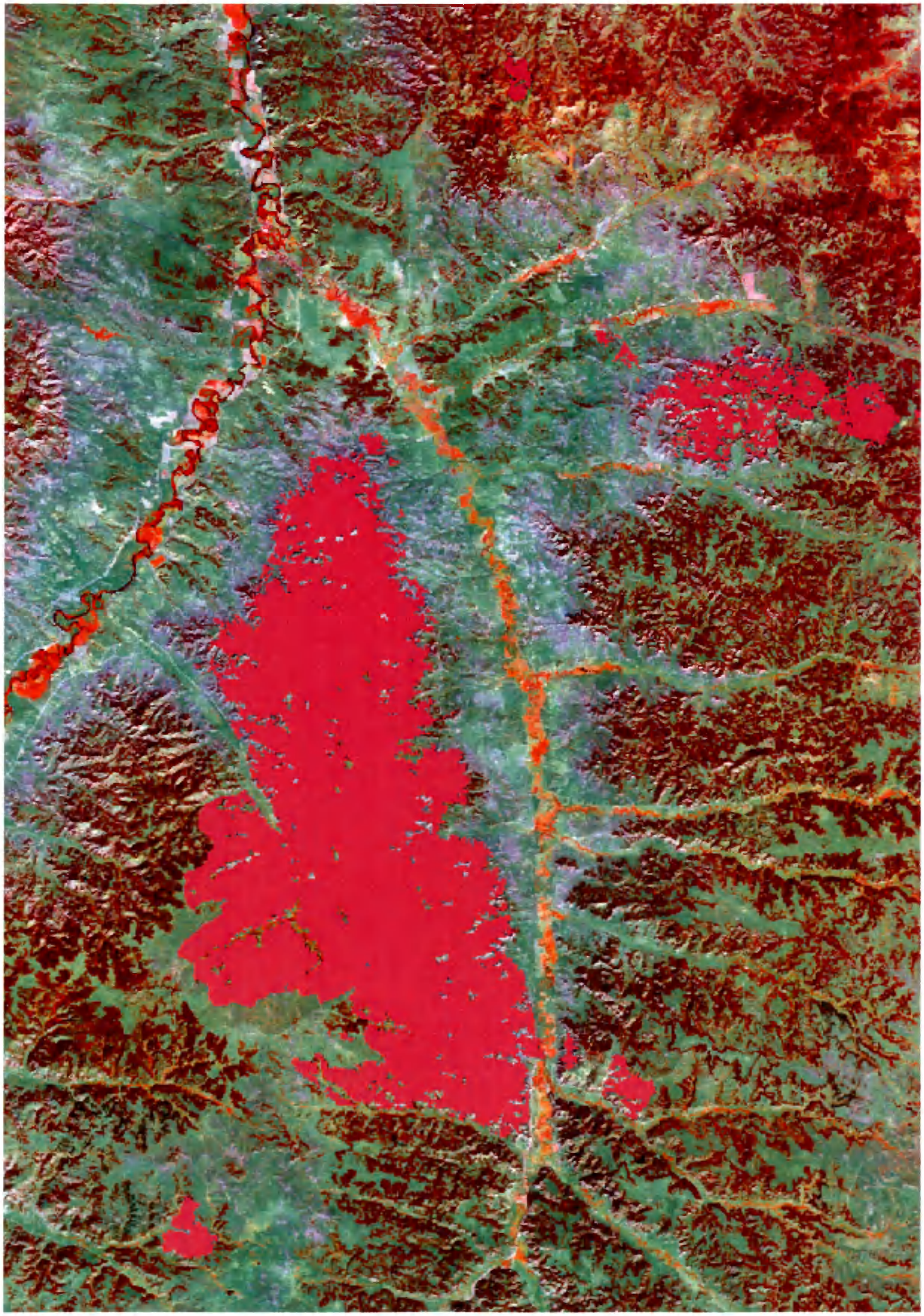
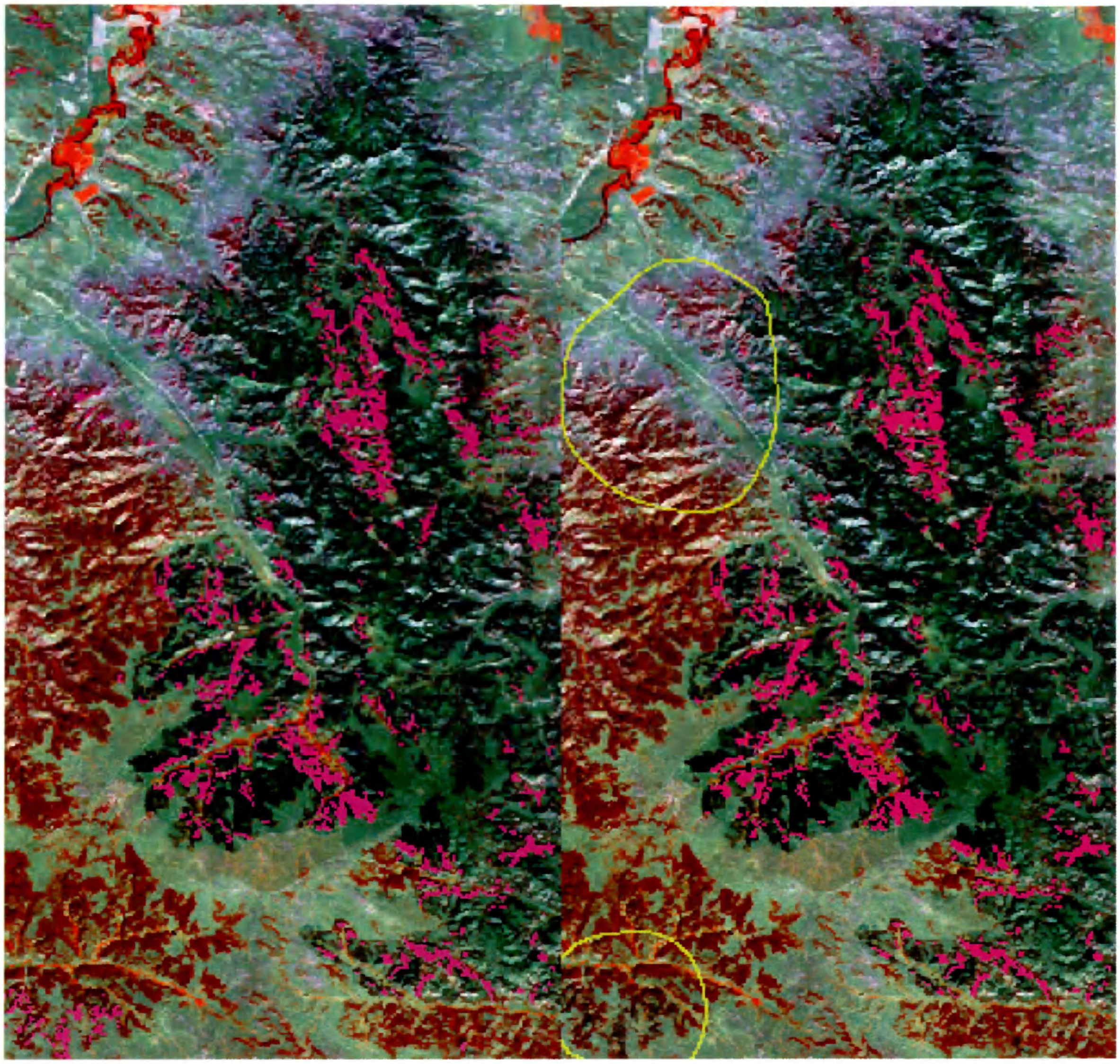


Figure 28: The Ashland burn perimeter delineated by the hierarchical feature extraction system.



(a) (b)

Figure 29: Clutter mitigation on "Mixed Burn" classification in the Ashland Fire Study Area. Notice Figure 29(a) contains mis-classified areas outside the burn perimeter. The areas circled in Figure 29(b) show the clutter mitigation results of Pass 2 filtering.

6. Future Work

This study opens up many avenues for future work. The technique of classifying an image with successive passes in a hierarchical fashion fosters ideas about other possibilities for using the hierarchical concept to apply passes that may specifically improve some aspect of the classification. Once Areas of Interest have been identified, we can apply numerous types of passes to improve results. They do not necessarily need to be learning classifiers. The following sub-sections briefly describe some novel ideas for improving classification robustness and accuracy.

6.1. Shape Matching

Shape matching algorithms fit a template form to the outline of an object's classification in an image. The idea is to morph the shape of the learner's classification to match the template. For example, notice the vehicles in the Presidio images are rectangular in shape, and the final machine classification identifies vehicles as blobs. The results would be more accurate and look better if they were rectangular in shape. This would work to mitigate clutter surrounding an object by cleaning up the learner's classification. Road and building classifications could also be improved by matching a template shape as a final pass in classification hierarchy.

6.2. *Rotational and Size Invariance*

A challenging aspect of classifying linear features such as roads or sidewalks is inferring an orientation that is not represented in the training examples. The problem is that when the examples only contain sidewalks oriented vertically but the image contains horizontal sidewalks the learner will often only correctly classify the horizontal sidewalks. Rotational or Size Invariance either rotates or scales the training examples to key the learner in on the fact that these features may be of different sizes or orientations. This technique could be applied as an early pass in the hierarchical process.

6.3. *Temporal Pattern Recognition*

Temporal pattern recognition allows the learner to identify a cartographic feature present in an image taken at any time of the year or time of the day. The basic idea is to create a learner that can make a prediction even though slight temporal changes may naturally occur within the scenes to be classified. The learner needs to be capable of recognizing temporal patterns as well as spatial and spectral patterns because, for example, the same feature may look entirely different in the summer than it does in the winter. A cartographic feature found in multiple images taken over time is a highly disjunctive concept. The learner must be capable of making a prediction even though the object may be able to change over time. Recognizing a feature in any weather condition and at any time of the day has many real applications, but it is a difficult task.

7. Conclusions

The goal of this study is to increase the accuracy of current image classification techniques through the use of a hierarchy of inductive learners. In this thesis we argue for the importance of extracting objects from digital images, outline reasons why the task is so difficult, and present a technique using a hierarchy of inductive learning algorithms to tackle the problem. Our approach is relatively simple, much quicker than traditional techniques, and improves the results of other automated feature extraction systems.

Semi-automated systems like the process described in this thesis must keep the amount of analyst input low, while using enough input and the right kind of information to guide the automated processes. The hierarchical system has this balance because it allows us the freedom to focus on areas where the classification task is difficult and to ignore areas that are easily recognized. The user quickly moves through the image to dynamically delineate areas of interest, simplifying the overall classification task. Although this technique is not a cut and dried solution, the hierarchical approach shows promise due to its ability to increase accuracy in iterative passes and narrow the classification task in the process. We show some exciting and positive results over current techniques more accurate, and can be faster than currently used manual and automated techniques.

8. References

Bain, W., (2000). MS Thesis, *A Comparison of Three Machine Learning Algorithms for Auotmated Feature Extraction from Digital Images.* Director: David W. Opitz, University of Montana.

Burl, M. et.al. (1998). Learning to Recognize Volcanoes on Venus, *Machine Learning Journal Special Issue on Applications and Knowledge Discovery.* 30:165-194.

Burl, M.C., U. Fayyad, P. Smyth, P. Perona and M.P. Burl (1994), Automating the hunt for volcanoes on Venus. In *Preceedings of the 1994 Conference on Computer Vision and Pattern Recognition*, 1994.

Ford, R., Ma, Z., Barsness, S. & Redmond, R. (1997). Rule-Based Aggregation for Region Identification. In *Preceedings of the 1997 American Society for Photogrammetry and Remote Sensing Annual Conference,* Seattle, WA.

Gruen, A., E.P. Baltsavias, O. Henricsson Birkhäuser Verlag Basel, Boston, Berlin, (1997) Automatic Extraction of Man-Made Objects from Aerial and Space Images (II) *Bi-annual Ascona conference, held in Ascona, Switzerland.*

Jensen, J. R. and D. C. Cowen, 1997, "Remote Sensing of Urban/Suburban Socioeconomic Attributes," *Land Satellite Information in the Next Decade II: Sources and Applications,* Bethesda: American Society for Photogrammetry & Remote Sensing, (December 5); CD, 19 p.

Kramber, W. J., and A. Morse. "Integrating Image Interpretation and Unsupervised Classification Procedures." *ACSM/ASPRS Annual Convention & Exposition.* Baltimore: ACSM/ASPRS, 1994. 1:327-336.

Langley, P., *Elements of Machine Learning*, Morgan Kaufman Publishers, Inc., 1996.

Lillesand, T., & Keifer,R., *Remote Sensing and Image Interpretation*, 4th ed: John Wiley & Sons, Inc., 1999.

Maloof, M. et al. (1998). Learning to detect rooftops in aerial images. In *Proceedings of the Image Understanding Workshop*, (pp. 835-845).

Mather, P. (1999). Computer Processing of Remotely-Sensed Images, An Introduction, 2nd ed.

Mitchell, T., *Machine Learning*, Boston: MIT Press, 1997.

Opitz, D., W. Bain, M. Mangrich, J. Zeiler, M. Sweet, (2000) *Rapid Extraction of Objects from Digital Images* Department of Computer Science, University of Montana

Palhang, M., A. Sowmya, (1997). Two Stage Learning, Two Stage Recognition, *Proceedings of Australian Joint Conference on Artificial Intelligence*, Perth Australia, 2-4 Dec. 1997, 1997.

Seul, M., L. O'Gorman, M. Sammon, (2000). *Practical Algorithms for Image Analysis: Description, Examples, and Code.*, Cambridge Univ Press.

Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285–1293.

Woodcock, C.E., and V.J. Harward, 1992. Nested-Hierarchical Scene Models and Image Segmentation, *International Journal of Remote Sensing*, 13(16):3167-3187.