Image Database Exploration: Progress and Challenges

Usama M. Fayyad and **Padhraic Smyth**
Mail Stop 525-3660
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109-8066
email: {*fayyad,pjs*}*@aig.jpl.nasa.gov*

Abstract

In areas as diverse as remote sensing, astronomy, and medical imaging, image acquisition technology has undergone tremendous improvements in recent years in terms of imaging resolution, hardware miniaturization, and computational speed. For example, current and future near-earth and planetary observation systems will return vast amounts of scientific data, a potential treasure-trove for scientific investigation and analysis. Unfortunately, advances in our ability to deal with this volume of data in an effective manner have not paralleled the hardware gains. While special-purpose tools for particular applications exist, there is a dearth of useful general-purpose software tools and algorithms which can assist a scientist in exploring large scientific image databases. At JPL we are currently developing interactive semi-automated image database exploration tools based on pattern recognition and machine learning technology. In this paper we discuss the general problem of automated image database exploration, the particular aspects of image databases which distinguish them from other databases, and how this impacts the application of off-the-shelf learning algorithms to problems of this nature. Current progress will be illustrated using two large-scale image exploration projects at JPL. The paper concludes with a discussion of current and future challenges.

1 Introduction

1.1 Background and Motivation

In a variety of scientific disciplines two-dimensional digital image data is now relied on as a basic component of routine scientific investigation. The proliferation of image acquisition hardware such as multi-spectral remote-sensing platforms, medical imaging sensors, and high-resolution cameras have led to the widespread use of image data in fields such as atmospheric studies, planetary geology, ecology, agriculture, glacielogy, forestry, astronomy, diagnostic medicine, to name but a few.

Across all of these disciplines is a common factor: the image data for each application, whether it be a Landsat image or an ultrasound scan, is but a means to an end in the sense that the investigator is only interested in using the image data to infer some conclusion about the physical properties of the target being imaged. In this sense, the image data serves as an intermediate representation to facilitate the scientific process of inferring a conclusion from the available evidence. This might seem like an obvious observation. Yet it could be argued that in many practical cases the process of acquiring and storing the images is seen as an end in itself and the subsequent image analysis is relegated to a minor role. Certainly this accusation could be made in the past of NASA's planetary science endeavours, where most of the resources were expended in the process of acquiring the data and relatively little consideration was given to how the data would actually be used after the mission was complete (e.g. see the 1992 Congress report on the Earth Observing System in this context [1]).

However, the climate of image acquisition and analysis is changing rapidly. In the past, in planetary science for example, image databases were analysed in a careful manual manner and much investigative work was carried out using hard copy photographs. However, due to the sheer enormity of the image databases currently being acquired, simple manual cataloging is no longer a practical consideration if all of the available data is to be utilised. As an example, consider the Magellan mission to Venus. The Magellan spacecraft transmitted back to earth a data set consisting of over 30,000 high resolution radar images of the Venusian surface. This data set is greater than that gathered by all previous planetary missions combined — planetary scientists are literally swamped by data. Venus is an extremely volcanic planet (volcanoes are by far the single most visible geologic feature in the Magellan data set), hence, the study of basic volcanic process is essential to a basic understanding of the geologic evolution of the planet [13]. Central to volcanic studies is the cataloging of each volcano location and its size and characteristics — there are estimated to be on the order of 10^6 visible volcanoes scattered throughout the 30,000 images [3]. It has been estimated that manually locating all of these volcanoes would require on the order of 10 man years of a planetary geologist's time to carry out — even then, a further search would be required to fully characterise the shape and appearance of each volcano. Given a catalog of volcanoes and their characteristics, a scientist can use the data to support various scientific theories and analyses. For example, the volcanic spatial clustering patterns may be correlated with other known and mapped geologic features such as mean planetary radius, which may provide evidence either pro or con particular theories of planetary history.

1.2 Scope and Outline

The Magellan-Venus data set is an example of a currently familiar pattern in the remote-sensing and astronomy communities — a new image data set becomes available but the size of the data set precludes the use of simple manual methods for exploration. Hence, scientists are beginning to express a need for automated tools which can assist them in navigating through large sets of images. A commonly expressed wish is the following: "is there a tool where I could just point at an object on the screen (or even draw a caricature of it) and then have the algorithm find similar items in the database?" Some scientists even have pre-conceived notions that neural networks or some other currently fashionable technology already provide a pre-packaged solution to their problem—we will argue that no such domain independent tools exist.

Note that in this paper the type of problem being addressed differs from the types of problems typically addressed by classical work in machine vision. Machine vision work has focused primarily on image understanding, parsing, and segmentation, with a particular emphasis on detecting and analysing *man-made* objects in the scene of interest. The focus of this paper is on the detection of *natural*, as opposed to *man-made*, objects. The distinction is important because, in the context of image analysis, natural objects tend to possess much greater variability in appearance than man-made objects. Hence, we shall focus primarily on the use of algorithms that "learn by example" as the basis for image exploration. The primary alternative, the model-based approach will not be dealt with except in passing. The "learn by example" approach is potentially more generally applicable since domain scientists find it relatively easier to provide examples of what they are searching for compared to describing a model. However, the distinction between prior models and "learning by example" should be viewed as two ends of a continuous spectrum rather than dichotomous points of view.

Using ongoing JPL projects as examples, the paper will examine the application of pattern recognition and machine learning technology to the general problem of image database exploration. In particular, it will be argued that image databases possess unique characteristics which impact the direct application of standard learning methods. Feature extraction from pixels, spatial context modeling, limited ground truth, the availability of prior knowledge, and the use of supervised feedback during learning are all common aspects of the problem which can either help or hinder the development of learning tools. Each of these issues will be discussed in the context of currently available learning theories and algorithms, and recent progress and opportunities for significant improvements will be outlined.

2 Two Illustrative Case Studies

To ground the discussion in this paper, we provide two illustrative examples of current projects at JPL involving the development of image exploration algorithms and tools. The first is an already successful application of decision tree learning to classification in the context of a well understood image analysis problem. The second project represents ongoing work which targets a more ambitious problem of dealing with domains where the basic image processing itself is not straightforward.

2.1 SKICAT: Automated Astronomy Sky Survey Cataloging

The first example consists of an application of machine learning techniques to the automation of the task of cataloging sky objects in digitized sky images. The Sky Image Classification and Archiving Tool (SKICAT) has been developed for use on the images resulting from the 2nd Palomar Observatory Sky Survey (POSS-II) conducted by the California Institute of Technology (Caltech). The photographic plates collected from the survey are being digitized at the Space Telescope Science Institute (STScI). This process will result in about 3,000 digital images of roughly 23,000×23,000 pixels[1] each. The survey consists of over 3 terabytes of data containing on the order of 10^7 galaxies, 10^8 stars, and 10^5 quasars.

The first step in analyzing the results of a sky survey is to identify, measure, and catalog the detected objects in the image into their respective classes. Once the objects have been classified, further scientific analysis can proceed. For example, the resulting catalog may be used to test models of the formation of large-scale structure in the universe, probe galactic structure from star counts, perform automatic identification of radio or infrared sources, and so forth. The task of reducing the images to catalog entries is a laborious time-consuming process. A manual approach to constructing the catalog implies that many scientists need to expend large amounts of time on a visually intensive task that may involve significant subjective judgment. The goal of our project is to automate the process, thus alleviating the burden of cataloging objects from the scientist and providing a more objective methodology for reducing the data sets. Another goal of this work is to classify objects whose intensity (isophotal magnitude) is too faint for recognition by inspection, hence requiring an automated classification procedure. Faint objects constitute the majority of objects on any given plate. We target the classification of objects that are at least one magnitude fainter than objects classified in previous surveys using comparable photographic material.

The learning algorithms used in SKICAT are the GID3* [9] and O-Btree [10] decision tree generation algorithms. In order to overcome limitations inherent in a decision tree approach, we use the RULER [11] system for deriving statistically cross-validated classification rules from multiple (typically > 10) decision trees. The details of the learning algorithms are beyond the scope of this paper and are therefore not covered here. For details of how rules are generated from multiple decision trees, and for comparisons with neural net performance, the reader is referred to [11]. Details of this problem are also covered in a companion paper in this proceedings [12].

[1] Each pixel consists of 16 bits and represents the intensity in one of three colors.

2.1.1 Attribute Measurement

A manual approach to classifying sky objects in the images is infeasible. Existing computational methods for processing the images will preclude the identification of the majority of objects in each image since they are at levels too faint (the resolution is too low) for traditional recognition algorithms or even methods based on manual inspection or analysis. Low-level image processing and object separation is performed by the public domain FOCAS image processing software developed at Bell Labs [18, 31]. In addition to detecting the objects in each image, FOCAS also produces basic attributes describing each object. These attributes are standard in the field of astronomy and represent commonly measured quantities such as area, magnitude, several statistical moments of core intensity, ellipticity, and so forth. Additional normalized attributes were measured later to achieve accuracy requirements and provide stable performance over different plates (see the discussion in Section 3.1). In total, 40 attributes are measured by SKICAT for each detected object.

2.1.2 Classifying Faint Objects and the Use of CCD Images

In addition to the scanned photographic plates, we have access to CCD images that span several tiny regions in some of the plates. The main advantage of a CCD image is higher resolution and signal-to-noise ratio at fainter levels. Hence, many of the objects that are too faint to be classified by inspection of a photographic plate, are easily classifiable in the corresponding CCD image (if available). We make use of the CCD images in two very important ways:

1. CCD images enable us to obtain class labels for faint objects in the photographic plates.
2. CCD images provide us with the means to reliably evaluate the accuracy of the classifiers obtained from the decision tree learning algorithms.

In order to produce a classifier that classifies faint objects correctly, the learning algorithm needs training data consisting of faint objects labeled with the appropriate class. The class label is therefore obtained by examining the CCD frames. Once trained on properly labeled objects, the learning algorithm produces a classifier that is capable of properly classifying objects based on the values of the attributes provided by FOCAS. Hence, in principle, the classifier will be able to classify objects in the photographic image that are simply too faint for an astronomer to classify by inspection of the survey images. Using the class labels, the learning algorithms are basically being used to solve the more difficult problem of separating the classes in the multi-dimensional space defined by the set of attributes derived via image processing. This method allows us to classify objects at least one magnitude fainter than objects classified in photographic sky surveys to date.

2.1.3 Results

It is important to point out that without the additional attributes described in Section 3.1, none of the learning methods achieved better than 75% accuracy. As expected, defining the new "normalized" attributes raised our performance on both intra- and inter-plate classification to acceptable levels varying between 92% and 98% accuracy with an average of 94%. Our encoding of these attributes represents an implicit imparting of more domain knowledge to the learning algorithm.

The SKICAT system is expected to speed up catalog generation by one to two orders of magnitude over traditional manual approaches to cataloging. This should significantly reduce the cost of cataloging survey images by the equivalent of tens of astronomer man-years. In addition, SKICAT classifies objects that are at least one magnitude fainter than objects cataloged in previous surveys. We have exceeded our initial accuracy target of 90%. This level of accuracy is required for the data to be useful in testing or refuting theories on the formation of large structure in the universe and on other phenomena of interest to astronomers.

The catalog generated by SKICAT will eventually contain about a billion entries representing hundreds of millions of sky objects. For the first survey (POSS-I) conducted over 4 decades ago which was without the availability of an automated tool like SKICAT, only a small percentage of the data was used and only specific areas of interest were studied. In contrast, we are targeting a comprehensive sky catalog that will be available on-line for the use of the scientific community. Because we can classify objects that are one magnitude fainter, the resulting catalog will be significantly richer in content, containing three times as many sky objects as would have been possible without using SKICAT.

As part of our plans for the future we plan to begin investigation of the applicability of unsupervised learning (clustering) techniques such as AUTOCLASS [5] to the problem of discovering clusters or groupings of interesting objects. The initial goals will be to answer the following two questions:

1. Are the classes of sky objects used currently by astronomers justified by the data: do they naturally arise in the data?
2. Are there other classes of objects that astronomers were not aware of because of the difficulty of dealing with high dimensional spaces defined by the various attributes? Essentially this is a discovery problem.

2.2 Volcano Detection in Magellan-Venus Data

The Magellan-Venus data set constitutes an example of the large volumes of data that today's instruments can collect, providing more detail of Venus than was previously available from Pioneer Venus, Venera 15/16, or ground-based radar observations put together [26]. We are initially targeting the automated detection of the "small-shield" volcanoes (less than 15km in diameter) that constitute the most abundant visible geologic feature [17] in the more that 30,000 synthetic aperture radar (SAR) images of the surface of Venus. It is estimated, based on extrapolating from previous studies and knowledge of the underlying geologic processes, that there should be on the order of 10^6 of these volcanoes visible in the Magellan data [3, 16].

Identifying and studying these volcanoes is fundamental to a proper understanding of the geologic evolution of Venus. However, locating and parameterizing them in a manual manner is forbiddingly time-consuming. Hence, we have undertaken the development of techniques to partially automate this task. The primary constraints for this particular problem are that the method must be reasonably robust and fast. Unlike most geological features, the small volcanoes can be ascribed to a basic process that produces features with a short list of readily defined characteristics differing significantly from other surface features on Venus [17]. For pattern recognition purposes the relevant criteria include (1) a circular planimetric outline, (2) known diameter frequency distribution from preliminary studies, (3) a limited number of basic morphological shapes, and (4) the common occurrence of a single, circular summit pit at the center of the edifice.

2.2.1 The Approach

There has been little prior work on detecting naturally occurring objects in remotely-sensed images. Most pattern recognition algorithms are geared towards detecting straight edges or large changes in texture or reflectivity. While this works well for detecting *man-made* objects, approaches such as edge detection and Hough transforms deal poorly with the variability and noise present in typical remotely sensed data [7, 22].

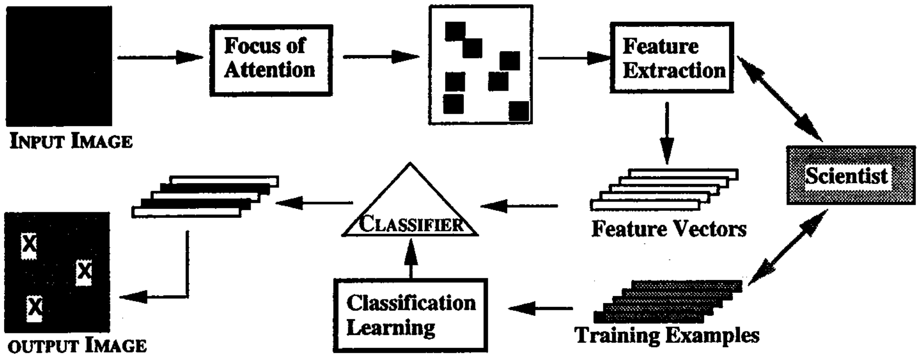

Figure 1: Block Diagram of the Proposed System

We are developing a system that consists of three distinct components: focus of attention, feature extraction, and classification learning. Figure 1 gives a block diagram of the approach.

Our initial work in this problem has relied on the concept of using a focus of attention (FOA) method to detect regions of interest followed by local classification of regions of interest into volcano and non-volcano categories. The focus of attention component is designed primarily for computational efficiency. Its function is to quickly scan an input image and roughly determine regions of interest (regions potentially containing objects similar to those specified by the scientist). For this purpose we have used a constant false alarm rate detector which compares the intensity of a center pixel with a locally adaptive estimate of the background intensity: if the central intensity is larger than some number of standard deviations from the background mean intensity, the region is considered detected. By running this detector at multiple resolutions of the image, it can detect both volcanoes at different scales and different features of the volcanoes [28]. For example, at high resolution it picks up the summit pit, while at lower resolutions the bright slopes can be detected. False alarms are caused by craters, grabens, and other bright features in the data.

Given a set of detected regions of interest, the remaining task is to discriminate between the volcanoes and false alarms. A current focus of the research is to find a useful feature-representation space — although nearest neighbour classifiers can provide reasonably accurate results (see section 2.2.2 below), a representation based purely on pixels will tend to generalize poorly. For the purposes of incorporating prior knowledge the ideal feature set would be expressed in the form of expected sizes, shapes, and relative geometry of slopes and pits, namely, the same features as used by the scientists to describe the volcanoes. However, due to the low signal-to-noise ratio of the image, it is quite difficult to gain accurate measurements of these features, effectively precluding their use at present. The current focus of our work is on a method which automatically derives robust feature representations — this will be described in Section 3.1.

2.2.2 Current Status and Preliminary Results

We have constructed several training sets using 75m/pixel resolution images labeled by the collaborating geologists at Brown University to get an initial estimate of the performance of the system. The FOA component, typically detects more than 80% of all the volcanoes, while generating 5-6 times as many false alarms. Using the nearest neighbour classifier, we can classify the regions of interest into volcanoes and false alarms with an estimated independent test accuracy of 82% — more recent results using features derived from both segmentation and principal component methods (see Section 3.1) has resulted in accuracies of the order of 85%. Similar accuracy results have been reported in [33] for this problem. It is important to clarify that these are initial results and with further effort we hope to be able to significantly improve the accuracy. Demonstrating the general applicability of this approach to the detection of other Venusian features as well as images from other missions will be the next step. So far the emphasis has been placed mainly on developing the computer tools to allow scientists to browse through images and produce training data sets (as well as partial catalogs) within a single integrated workstation environment.

3 The Role of Prior Information

In general, prior information can be specified in two ways. The first is in terms of relatively high-level knowledge specifying expectations and constraints regarding certain characteristics of the objects of interest. For example, in the Magellan-Venus problem the incidence angle of the synthetic aperture radar instrument to the planet's surface is known, which in turn strongly influences the relative positions of bright and dark slope and summit regions for a given volcano [21].

The second type of prior information which we consider here is normally not thought of as such. This is the information which is implicitly specified by the labeled data, i.e., the data which has been examined by the domain expert and annotated in some manner. While one normally thinks of the labeled data and the prior knowledge as two separate entities, it is convenient in practice to consider both the knowledge and data forms of prior information within the same context.

One must determine the degree of utility of each type of information in designing an exploration algorithm. For example, in the SKICAT project, the prior knowledge was quite precise and helped a great deal in terms of determining the optimal features to use for the problem. In contrast, for the Magellan-Venus problem, the prior knowledge is quite general in nature and is not easily translatable into algorithmic constraints. Hence, thus far, the most effective source of prior information has been the labeled training examples provided by the scientists.

Below we consider two important aspects of prior information. The first addresses the issue of deriving suitable higher-level representations from the raw pixels. The second issue concerns the nature of the labeled data provided by the domain expert.

3.1 Pixel Data versus Feature Data

Raw pixel data is rarely useful or of interest to users. Humans typically perform some sort of pixel-to-feature mapping immediately. In scientific data analysis domains, where the user typically knows the data well and has a list of defined features, using this knowledge makes the learning task significantly easier. SKICAT provides an excellent example of this. Not only was the segmentation problem (locating objects) easy to perform, but we had access to a host of defined attributes that we made use of effectively. Having the proper representation made the difference between success and failure in that case.

In order for SKICAT to achieve stable classification accuracy results on classifying data from different plates, we had to spend some effort defining normalized attributes that are less sensitive to plate-to-plate variation. These attributes are computed automatically from the data, and are defined such that their values would be normalized across images and plates. Many of these quantities (although not all) have physical interpretations. Other quantities we measured involved fitting a template to a set of "sure-stars" selected by the astronomer for each image, and then measuring the rest of the objects with respect to this template. In order to automate the measurement of such attributes, we automated the "sure-star" selection problem by treating it as a learning sub-problem and building decision trees for selecting "sure-stars" in an arbitrary image[2]. It is beyond the scope

[2]This turns out to be a relatively easy learning task, our accuracy on this subproblem exceeds 98%.

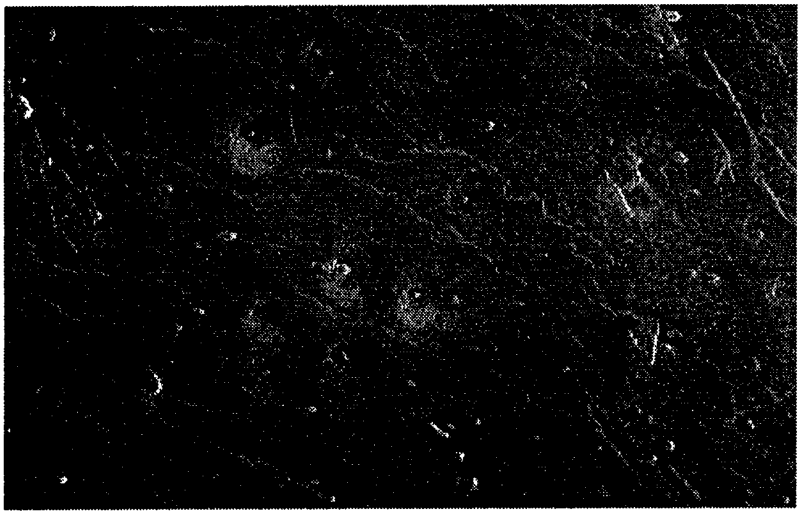

Figure 2: An Example Magellan Subimage of Venus Illustrating Small Volcanoes.

of this paper to give the detailed definitions of the attributes. The point is that in this case a wealth of knowledge was available to us in terms of attributes (measurements), while astronomers had little knowledge of *how to use* these attributes to classify objects.

On the other hand, in the case of Magellan SAR images, the image segmentation problem (object detection) is significantly more difficult to address. In general, one needs to address the problem of automatic feature (attribute) construction. One approach we have been experimenting with for this purpose is the use of principal component analysis. Each training example (subimage containing positive example) can be turned into a vector of pixel values. The entire training set will thus form a $k^2 \times n$ matrix[3] which can subsequently be decomposed into a set of orthonormal eigenvectors using singular value decomposition (SVD). An eigenvalue is associated with each of the vectors indicating its relative importance. When the eigenvectors (eigenvolcanoes) are viewed as images again, we note that each represents a "basic" feature of a volcano. Figure 2 shows an example Magellan Venus image with a few small volcanoes showing. Figure 3 shows 15 associated eigenvolcano features (those corresponding to the largest eigenvectors) ordered left to right by decreasing eigenvalues. Note that the eigenvectors become less coherent starting with the sixth or seventh feature. Each block in the figure corresponds to a 225-component eigenvector that was re-translated into a 15x15 image and redisplayed as a block in the image.

The eigenvolcanoes can be viewed as general features that can be used to encode each detected candidate volcano for classification purposes. This is an example of an automatic template (matched filter) generation procedure which can easily be augmented by other features provided by the expert user.

3.2 Supervised Feedback and the Lack of Ground Truth

It is commonly assumed in learning and pattern recognition algorithms that the categorical class labels attached to the training data represent ground truth. In fact, it is often the case that this is not so and that the class labels are subjective estimates of ground truth as provided by an expert. The distinction is an important one. In particular, the question arises as to whether or not the

[3] Assume that there are n examples, each of which consists of a $k \times k$ pixel subimage.

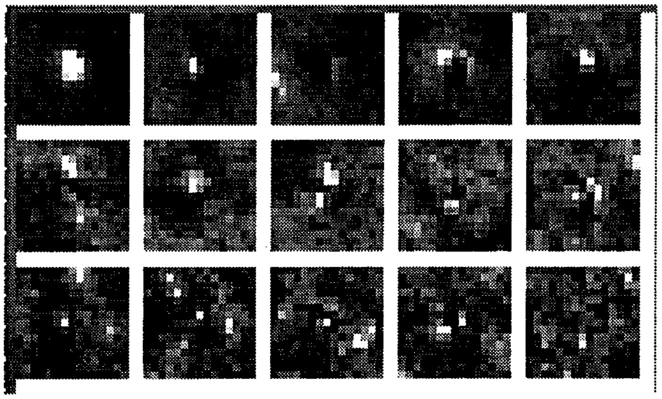

Figure 3: Eigenvolcanoes Derived from Training Data.

expert should provide his or her best guess at the class label of each exemplar, or instead should provide some quantitative estimate of the likelihood that the exemplar is a member of some class. Smyth [29] has shown that virtually all well-known learning algorithms can be easily modified to handle such probabilistic labels if one assumes that the expert is providing true unbiased estimates of the likelihoods. In practice, however such subjective estimates are likely to be biased and inconsistent. It is quite difficult to accurately elicit subjective probabilities from even the most cooperative human subject — this has been well documented in various studies which have shown that even domain experts are typically quite inconsistent in their estimates of their own subjective probabilities [6].

The volcanoes in the Magellan-Venus data can be quite ambiguous in appearance. Initially domain experts provided "hard" decisions on each example. This proved unsatisfactory since it ignored the obvious uncertainty in the more ambiguous cases. The current method of choice is to use *quantised* probability bins, three in all, each corresponding to a clearly defined and commonly agreed upon interpretation. Both for model development and for accurate model calibration, even the simple quantised probabilistic estimates have been a significant improvement over the hard decision labels. For example, the quantised labels are the basis for a more accurate loss function which is quite useful in terms of evaluating algorithm performance; algorithm errors on the more ambiguous volcanoes incur a lower loss function penalty than errors on the "certain" ones.

A further complication which can arise in practice is that of multiple experts. This can mean that each training example is subject to a different interpretation by different experts. There is a considerable literature on various methods which combine the beliefs of different individuals [15] — however most of these methods are of theoretical interest only since they assume that one has an excellent model of the correlation between the experts, something which is difficult to estimate in practice. Simple linear combination techniques should give reasonable results in most cases. In practice, for the Magellan-Venus data, having the experts cooperate to produce a consensus estimate for each example seems to work well.

4 Detection & Classification: Learning from Positive Examples

A significant challenge in dealing with image databases is the management of computation over a very large number of pixels. For building an analysis tool that is to be used in real-time one cannot afford to apply expensive measurement routines indiscriminantly. From a practical point of view this means that a focus of attention (FOA) method needs to be applied. Since the FOA needs to be efficient, it may generate many false alarms. The task of distinguishing truth from false alarms falls on the learning algorithm. This automatically brings in the issue of whether one is learning from only positive or from both positive and negative examples.

The user is interested in providing only positive examples. One can choose to learn only from these examples. However, extra information can be exploited for free by allowing the learning algorithm to learn and exploit the biases of the FOA. By applying the FOA on the images from which training data was obtained, a training set of positive and negative examples can easily be constructed. This approach, however, requires that the expert did not miss any positive examples in the labeled image, since any objects missed by the expert are likely to be picked up by the focus of attention detector and will be incorrectly interpreted as false alarms (this has occurred in practice in the volcano problem where some images may contain on the order of 100 volcanoes). The other disadvantage is that "negative" examples are not really meaningful to the user. Hence, the learned classifier will not necessarily be interpretable by the user since it is discriminating between true examples and by-products of the FOA. The point is that negative examples arise naturally as a consequence of computational efficiency considerations and their use may significantly facilitate the discrimination learning task.

Another choice is to perform detection and classification in one step. In problems where a comprehensive careful analysis is the target (as in SKICAT), one can perform sophisticated expensive detection and measurement on all pixels in the data. This means that the system can no longer respond in a short amount of time and also implies that methods for performing segmentation in a reliable manner have been developed. In the case of the Magellan SAR images of Venus, no general off-the-shelf algorithms are available. However, this does not rule out the use of matched filters (templates) to perform both detection and classification in one step (assuming the required computation is practical). Decomposing the problem into an FOA stage followed by a classification stage generally makes each sub-problem easier to tackle and solve.

5 Modelling Spatial Context

The two-dimensional nature of spatial data means that pixel elements in an image database are likely to be highly correlated. Most discrimination and classification algorithms implicitly assume that the training data they are dealing with consists of independent randomly chosen samples from the population of interest, e.g., a set of medical records for a hospital. Hence, in theory, they are not directly applicable to the problem of learning pixel classification maps (for example). Nonetheless, much of the work in remote sensing until recently has focused on local pixel classification methods, whereby the estimated categorical label of a pixel is purely a function of the intensity of that pixel and independent of the properties of neighbouring pixels [23]. This purely local estimation is non-intuitive and does not accurately reflect the human visual process whereby prior expectations and constraints are imposed so that global spatial coherence is obtained in the final labelling. To solve this, one can impose spatial smoothness constraints on both the labels and the pixel intensities. The most advanced such models were developed for practical applications in the mid to late 1980's under the general framework of Markov Random Fields (MRF's) [14, 25]. While the theoretical basis of MRF's is quite solid it is important to remember that they are primarily used as a computational convenience rather than a realistic model of spatial interaction. Other, more global, models of spatial context have also been proposed [2, 19], again with a sound mathematical basis.

However, it is fair to say that much work remains in terms of improving image spatial models.

There is a lack of theory on how to specify spatial models (such as MRF's) from prior knowledge. In particular, the parameters of the various MRF approaches must be set by the user and can be quite non-intuitive — in fact, these parameters often appear to be chosen in an ad hoc manner. Yet setting the parameters *a priori* is currently the only viable approach, since it is not possible to learn MRF's from data because of their non-causal nature except in special circumstances [8]. Hence, from an algorithm designer's viewpoint the situation is less than ideal when it comes to modelling spatial interaction — it appears that considerable experimentation and tuning is often necessary to find the right model for a given application.

6 Online Learning and Adaptation

Another aspect of the image exploration problem is that one would ideally like to have an algorithm which could gradually improve its performance as it explores more and more of the database. In fact this type of *incremental adaptation* is a desirable feature in many practical applications but has largely been ignored by researchers in learning and pattern recognition in favour of the simpler problem of "one-shot" batch learning. The model representation being used critically influences whether the model is easily adaptable or not. Discriminative models which focus on the differences between categories typically have trouble adjusting to new data in an elegant manner — it may be possible to easily adapt the parameters of the model but not the structure (consider decision trees as an example). Memory and prototype-based models (including parametric densities, non-parametric density estimators, mixture models, nearest-neighbour models, etc.) are naturally more suited to online adaptation — however, they typically suffer from poor approximation properties in high dimensions [27]. Hybrid models which combine the better features of discriminative and memory models would appear to have promise, however, there has been little work in this area.

In practice, an online image exploration algorithm would work by iterative interaction with the human user. The human visual system of the domain expert offers an excellent opportunity for supervised feedback to improve adaptation. This is in contrast to typical learning applications from "flat" data where there is no obvious intuitive way for a human labeller to visualize high-dimensional vectors. Hence, a reasonable strategy is to have the algorithm periodically query the domain expert for feedback on particular examples. In a probabilistic context it can be shown that the most information can be gained by queries about examples which are in the areas of greatest posterior uncertainty — an algorithm can learn the most by getting feedback on the examples it is most unsure of. This has the effect of making the most efficient use of the queries — a "blind" algorithm which produced random examples for supervised feedback would quickly exhaust the patience of any human observer. Given unlabelled examples, the algorithm can perform unsupervised adaptation such as "decision-directed" learning where the algorithm uses its current model to label a new example and then updates its model as if that were the correct decision (such methods have been well-studied in the adaptive signal processing literature with applications to problems such as channel equalization). This can be effective in speeding convergence once an initially good model is obtained but can obviously diverge from the ideal solution if the model is inaccurate to begin with.

Yet another useful application of the online adaptation idea is the notion of selective model refinement, i.e., allowing the user to tune the detection model from a general to a more specific model. For example, in the Magellan-Venus database, there are many subclasses of volcanoes within the general class. Ideally, the planetary scientists would like to be able to modify the volcano detection model in order to restrict the search to specific types of volcano, based on appearance or size. The preference can be stated explicitly in the form of high-level constraints ("only consider volcanoes of diameter less than 3km") or can be implicitly provided in the form of examples of the specialised concept. Once again the type of model being used critically influences the manner by which it can be refined. For example, models which use an explicit knowledge representation such as decision rules can easily incorporate explicitly-specified constraints provided the language of representation is well-matched. Implicit discrimination models, such as neural networks, are better suited to dealing with implicit constraints in the form of data than explicit constraints, and can use the new data to project the existing model into subspaces of the existing decision regions.

7 Multi-Sensor and Derived Map Data

It is relatively common in remote-sensing applications to illuminate the target at multiple wavelengths, thus obtaining a vector of intensities at each pixel site rather than just a single intensity. In the Magellan-Venus data for example, many parts of the planet were imaged from different angles and at different resolutions, resulting in several different data sets being available for the same surface regions. Low-resolution altimeter data was also measured providing a low-resolution map of the mean planetary radius.

Similarly, after data has been acquired and archived, different research groups will typically analyse the data and produce thematic maps and catalogs (either by manual or automated means) for different quantities of interest [4, 20]. For example, in the Magellan-Venus database, catalogs have already been produced for large volcanic structures and for the location of many of the large volcanic fields (but not the volcanoes within the fields).

Hence, in the general sense, each pixel can have a vector of associated attributes, whether these are data from another sensor, or derived qualitative categories (such as a map). In principle, such additional data should be particularly useful for computer-aided detection since it is often difficult for a human user to visualize such multi-dimensional representations. However, certain technical difficulties must be overcome for the additional data to be useful. For multi-sensor data, the different data sets must usually be *registered* so that the pixel measurements are somehow aligned to reference the same surface point — inevitably this is an imprecise process and spatial errors result. For qualitative map data one would like to ascertain the reliability of the map categories. It would be extremely useful if the map data contained not only the category label but also the degree of confidence ("spatial error bars") in that labelling. This is not done in subjective manual mapping for the obvious reason that the elicitation of such error bars would be a tedious and inaccurate process. However, automated map-making tools in general should provide some self-calibrated estimate of the reliability of the decision at each pixel or region of interest — algorithms based on probabilistic models (such as Bayesian methods) automatically provide such information.

8 Conclusion

Natural object detection and characterization in large image databases is a generic task which poses many challenges to current pattern recognition and machine learning methods. This paper has briefly touched on a number of relevant issues in problems of this nature: prior information, deriving features from pixel data, subjective labelling,, learning from positive examples, models for spatial context, online learning, and multi-sensor and thematic data. There are other issues which were not discussed here due to space constraints: the use of physical noise models for the radar imaging processes and other non-visible wavelengths, the integration of multiple images of the same surface area taken at different times, and the use of multi-resolution and parallel algorithms to speed computation.

The SKICAT and Magellan SAR projects are typical examples of the types of large-scale image database applications which will become increasingly common — for example, the NASA Earth Observing System Synthetic Aperture Radar (EOS SAR) satellite will generate on the order of 50 GBytes of remote sensing data per hour when operational [32]. In order for scientists to be able to effectively utilise these extremely large amounts of data, basic image database navigation tools will be essential.

Our existing JPL projects have so far demonstrated that efficient and accurate tools for natural object detection are a realistic goal provided there is strong prior knowledge about how pixels can be turned into features and from there to class categories. With the astronomy problem there was sufficient strong knowledge for this to be the case: with the volcano data, the knowledge is much less precise and consequently the design of effective object detection tools is considerably more difficult.

The common thread across the various issues would appear to be the problem of how to combine both prior knowledge and data. Much of the prior knowledge of a domain scientist is vague and imprecise and cannot be translated easily into pixel-level constraints. However, scientists find it significantly easier to provide attributes to measure on a given region than to specify the method they use to classify the region.

Dealing with image data is uniquely appropriate for interactive tools since results can immediately be visualized and judged by inspection. This makes obtaining feedback and training data from users much easier. Since humans find it particularly difficult to express *how* they perform visual detection and classification, using a "learning from examples" approach becomes particularly appropriate. The fact that the image databases are becoming increasingly common and unmanageably large makes the need for the type of approaches advocated in this paper particularly pressing.

Acknowledgments

We would like to acknowledge our collaborators on SKICAT at Caltech's Astronomy Department: N. Weir and S.G. Djorgovski. The Magellan SAR work is in collaboration with M.C. Burl and P. Perona, E.E. Department, Caltech; the domain scientists are J. Aubele and L. Crumpler, Dept. of Geological Sciences, Brown University. The research described in this paper was carried out in part by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] "Earth Observing System—NASA's EOSDIS Development Approach is Risky", *GAO [General Accounting Office] Report: GAO/IMTEC-92-24*, Feb. 1992.

[2] Y. Amit, U. Grenander, and M. Piccioni, "Structural image restoration through deformable templates," *J. American Statistical Association*, 86(414), pp.376–387, June 1991.

[3] J. C. Aubelle and E. N. Slyuta, "Small domes on Venus: characteristics and origins," in *Earth, Moon and Planets*, **50/51**, 493–532, 1990.

[4] P. A. Burrough, *Principles of Geographic Information Systems for Land Resources Assessment*, Oxford: Clarenden, 1986.

[5] P. Cheeseman, M. Self, J. Kelly, W. Taylor, D. Freeman, and J. Stutz, (1988) "Bayesian Classification." *Proc. of the Seventh National Conf. on Artificial Intelligence AAAI-88*, pp. 607-6611, Saint Paul, MN.

[6] R. M. Cooke, *Experts in Uncertainty: Opinion and Subjective Probability in Science*, Oxford University Press, New York, 1991.

[7] A.M. Cross, *Int. J. Remote Sensing*, 9,no.9, 1519-1528, 1988.

[8] P. A. Devijver and M. M. Deksel (1987) "Learning the parameters of a hidden Markov random field image model: a simple example," in *Pattern Recognition Theory and Applications*, P. A. Devijver and J. Kittler (eds.), Berlin: Springer-Verlag, pp.141–164.

[9] U.M. Fayyad (1991). *On the Induction of Decision Trees for Multiple Concept Learning.* PhD dissertation, EECS Dept., The University of Michigan.

[10] Fayyad, U.M. and Irani, K.B. (1992). "The attribute selection problem in decision tree generation" *Proc. of the Tenth National Conference on Artificial Intelligence AAAI-90* (pp. 104-110). Cambridge, MA: MIT Press.

[11] U. Fayyad, N. Weir, and S.G. Djorgovski (1993). "SKICAT: a machine learning system for automated cataloging of large scale sky surveys." *Proc. of Tenth Int. Conf. on Machine Learning*, Amherst, MA.

[12] U. Fayyad, N. Weir, and S.G. Djorgovski (1993). "Automated analysis of a large-scale sky survey: The SKICAT system." *Proc. of the Knowledge Discovery in Databases Workshop, AAAI-93*, San Mateo, CA: AAAI Press.

[13] J. W. Head et al. (1991) "Venus volcanic centers and their environmental settings: recent data from Magellan," EOS 72, p.175, American Geophysical Union Spring meeting abstracts.

[14] S. Geman and D. Geman (1984) "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Mach. Int.*, vol.6, no.6, 721–741.

[15] C. Genest and J. V. Zidek (1986) "Combining probability distributions: a critique and an annotated bibliography," *Statistical Science*, vol.1, no.1, pp.114–118.

[16] J. W. Head et al. (1992) *Journal Geophysical Res.*, 97, E8, 13,153-13,197.

[17] J. E. Guest et al. (1992) *Journal Geophys. Res.*, 97, E10, 15949.

[18] J. Jarvis and A. Tyson (1981). *Astronomical Journal* 86:41.

[19] M. I. Miller, G. E. Christensen, Y. Amit (1993) "A mathematical textbook of deformable neuroanatomies,' submitted to *Science.*

[20] D. F. Marble and D. J. Peuquet (1983) "Geographical information systems and remote sensing," in *Manual of Remote Sensing*, 2nd ed., R. E. Colwell (ed.), Falls Church, VA: Amer. Soc. Photogrammetry.

[21] G. H. Pettengill et al. (1991) "Magellan: radar performance and products," *Science*, vol.252, 260–265, 12 April 1991.

[22] S. Quegan et al, (1988) *Trans. R. Soc. London*, A 324, 409-421.

[23] J. A. Richards (1986) *Remote Sensing for Digital Image Analysis*, Springer-Verlag, Berlin.

[24] E.Rignot and R. Chellappa (1992) "Unsupervised segmentation of polarimetric SAR data using the covariance matrix," *IEEE Trans. on Geoscience and Remote Sensing,* vol.30, no.4, pp.697-705.

[25] B. D. Ripley (1988) *Statistical Inference for Spatial Processes*, Cambridge University Press, Cambridge.

[26] *Science, special issue on Magellan data*, April 12, 1991.

[27] P. Smyth and J. Mellstrom (1992) "Detecting novel classes with applications to fault diagnosis," in *Proceedings of the Ninth International Conference on Machine Learning*, Morgan Kaufmann Publishers: Los Altos, CA, pp.416–425.

[28] P. Smyth et al, (1992) "Multi-Resolution Pattern Recognition of Small Volcanoes in Magellan Data", *Proc. of the Int. Venus Colloq.*, California Institute of Technology, August 1992.

[29] P. Smyth (1993) "Learning with stochastic supervision," in *Computational Learning Theory and Natural Learning Systems 3*, T. Petcshe, M. Kearns, S. Hanson, R. Rivest (eds), Cambridge, MA: MIT Press, to appear.

[30] P. H. Swain, S. B. Vardeman, and J. C. Tilton (1981) "Contextual classification of contextual data," *Pattern Recognition*, 13, pp.429–441.

[31] Valdes (1982). Instrumentation in Astronomy IV, SPIE vol. 331, no. 465.

[32] J. Way and E. A. Smith (1991) "The evolution of synthetic aperture radar systems and their progression to the EOS SAR," *IEEE Trans. on Geoscience and Remote Sensing*, vol.29, no.6, pp.962–985.

[33] C.R. Wiles and M.R.B. Foreshaw (1992) *Lunar & Planetary Sciences Conference XXIII* (abst), 1527-1528.