Proceedings of the 2001 IEEE
International Conference on Robotics & Automation
Seoul, Korea - May 21-26, 2001

Autonomous Characterization of Unknown Environments

Liam Pedersen*
The Robotics Institute
Carnegie Mellon University
Pittsburgh PA 15213, USA

Abstract

Key to the autonomous exploration of an unknown area, by a scientific robotic rover is the ability of the vehicle to autonomously recognize objects of interest and generalize about the region. This paper presents a Bayesian framework under which a mobile robot can learn how different classes of objects are distributed over a geographical region, using imperfect observations and non-random sampling. This yields dramatic improvements in classification accuracy by exploiting the interdependencies between objects in an area and allows the robot to autonomously characterize the region. This is demonstrated with data from Carnegie Mellon University's Nomad robot in Antarctica, where it traversed the ice sheet, classifying rocks in its path.

Introduction

For many tasks involving the robotic searching and exploration of an area to find and identify objects, it is necessary to characterize the operating environment. To classify objects using onboard sensors, the likely candidates and potential false targets must be known in advance. In addition, to optimally classify objects their prior probabilities of being encountered must be known.

Unfortunately, it is difficult to know *a priori* the relative chances of finding different objects in an unexplored area. Consider the problem of classifying rocks from a robot for the purpose of geological exploration. There are many possible rock types, some hard to distinguish from each other. A geological map, if available, only indicates the most common rock type over a very large area. It does not indicate all rock types or their relative probabilities, and ignores small-scale variations. The latter is very important, as the kinds of rocks present can change significantly over a short distance, such as when crossing geological strata.

The standard approach to classifying objects is to consider each one independently, and classify it based on observations. However, objects in an area may be correlated, exploiting this can significantly increase classification accuracy. It is common for objects of the same type to be clustered together.

This paper will show a Bayesian approach to using the dependencies between objects distributed over an area by learning the statistical prior probabilities for different objects as a function of position, and specifically from the perspective of exploring an area with a mobile robot. Classification is thus improved not only by exploiting spatial dependencies, but also through improved knowledge of the priors. Furthermore, a map of prior probabilities over a geographic area is itself a useful summary with which to characterize the region and recognize gross properties.

This work is applicable to a variety of tasks involving the classification of objects distributed across an area, such as geological exploration, landmine removal, or soil profiling for agriculture. However, this analysis focuses on the problem of identifying the abundances of different rock types in an area of the Antarctic ice sheets using a scientific robotic vehicle, Nomad (Figure 1), built at Carnegie Mellon University to look for meteorites in Antarctica [1].

Figure 1 *Nomad robot in Antarctica, investigating rocks with a spectroscopic sensor to classify them.*

Autonomously identifying rocks with a robot is a challenging task with a high error rate [2]. It is highly relevant to the next generation of planetary rovers for exploring Mars, intelligently selecting samples for return to Earth, and incorporates the issues that arise when classifying objects from a mobile robot.

Robotic learning of the environment

Using a robotic vehicle to explore an area introduces unique issues.

- Sampling is not random. Rocks are examined by the robot as it traverses a path through the terrain of interest, leaving areas unexplored (Figure 2). The uncertainty about what is found in these areas must be noted, and

* Email: pedersen+@ri.cmu.edu

used to constrain subsequent changes in beliefs about the area as new data is added.

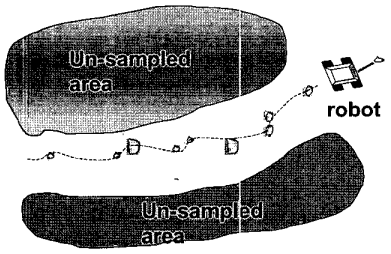

Figure 2 *Modus operandi of a robotic explorer. Only selected samples are visited in a large area, and they are not randomly distributed, but lie along the robots path.*

- Relatively few rock samples are examined. The Nomad robot (Figure 1) obtained measurements of no more than 50 rocks in 2 days. To make matters worse, there are many possible rock classes [3]. The rock probabilities must be therefore be initially coarsely defined and subsequently improved if and when data becomes available.

- Rock samples cannot usually be identified with complete certainty. Rather, when sensor data is obtained from a rock sample, only the likelihoods of different rock classes generating that data is known.

- The probabilities of different rock types are conditioned on geographical position, a continuous 2D (or 3D) quantity.

While machine learning and statistical estimation are mature fields, little prior work directly addresses the problem of characterizing a geographical area for the purposes of classification. The evidence grids of [4] are used to model the likelihoods of obstacles in an area. However, they fail to account for statistical dependencies between objects and require that space be discretized into a grid.

[5] and [6] survey strategies for either autonomously searching an area, given a prior description of the geographical distribution of targets, or exploiting the knowledge that targets tend to cluster together. They do not address how this information is obtained in the first place.

Representing the rock priors over an area

Consider the parameter $\vec{\theta}_x = [\theta_{x1},...,\theta_{xM}]$, representing the relative proportions of each rock type present at a geographic location x, and the random variable $\mathbf{R_x}$ the rock type (class labels 1, 2... M) of a rock sample found at there. Therefore

$$\theta_{x1},...,\theta_{xM} \in [0,1],\ \sum_{i=1}^{M}\theta_{xi} = 1 \quad (1)$$

$$P(\mathbf{R_x} = \mathbf{k} \mid \vec{\theta}_x) = P(\mathbf{R_x} = \mathbf{k} \mid \theta_{xk}) = \theta_{xk} \quad (2)$$

$$P(\mathbf{R_x} = \mathbf{k}) = E\{\theta_{xk}\} \quad (3)$$

$\vec{\theta}_x$ is itself a random variable and depends on position. Furthermore, knowing its distribution allows the determination of the rock type priors at x (3). Therefore, the problem of learning these priors is solved by learning the distribution of $\vec{\theta}_y$ at all positions y, given a sequence of robot sensor observations $\{\mathbf{O}_{x_i} | i=1,2...N\}$ made at locations $\{x_i \mid i = 1,2...N\}$. To do this it is necessary to model the statistical relationships between $\vec{\theta}_y$, the rock types of each rock sample, and the associated observations. Consequently we will show how to compute $p(\vec{\theta}_y | \mathbf{O}_{x_i})$, the posterior representation of the rock priors over the area, and how to use it to improve classification.

Pseudo Bayes network generative model

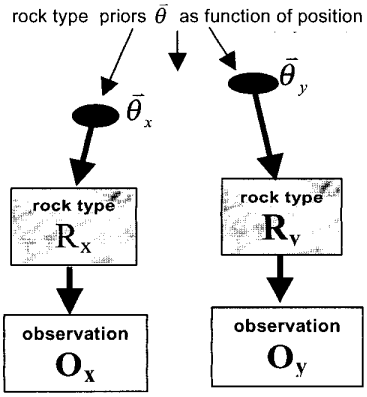

Figure 3 *Pseudo Bayes network generative model of the statistical relationships between position (x,y), relative proportions of different rock types at those positions* $(\vec{\theta}_x, \vec{\theta}_y)$ *, type of rock samples found* (R_x, R_y)*, and the observations on those samples* (O_x, O_y)*.*

It is reasonable to assume that the rock type of a sample at any given location x is conditionally dependent only on the local rock ratios $\vec{\theta}_x$, and sensor observations of a rock depend only upon its type (Figure 3). Therefore, for any positions x,y, and rock type $k \in \{1,2,...,M\}$

$$P(R_x = k \mid \vec{\theta}_x, \vec{\theta}_y, R_y) = P(R_x = k \mid \vec{\theta}_x) = \theta_{xk} \quad (4)$$

No assumptions are yet made on the relationship between $\vec{\theta}_x$ and $\vec{\theta}_y$, which is not usefully expressible in terms of a standard Bayes network diagram.

Furthermore, (4) implies that

$$\begin{aligned} & P(R_x = k \mid \vec{\theta}_y) \quad (5) \\ &= \iint P(R_x = k \mid \vec{\theta}_x, \vec{\theta}_y) p(\vec{\theta}_x \mid \vec{\theta}_y) d\vec{\theta}_x \\ &= \int \theta_{xk} \left\{ \int p(\vec{\theta}_x \mid \vec{\theta}_y) d\theta_{x1} .. d\theta_{xk-1}, d\theta_{xk+1} .. d\theta_{xM} \right\} d\theta_{xk} \\ &= \int \theta_{xk} p(\theta_k \mid \vec{\theta}_y) d\theta_{xk} \\ &= \mathrm{E}\{\theta_{xk} \mid \vec{\theta}_y\} \end{aligned}$$

a result that will be used subsequently. See [7] for more on the Bayes network representation of statistical relationships.

Geographical models

Consider the case when rocks can be identified with complete certainty. Suppose a rock at position x is determined to be of type k. Then, using Bayes rule, the posterior density of $\vec{\theta}_y$ is given by

$$p(\vec{\theta}_y \mid R_x = k) = \frac{P(R_x = k \mid \vec{\theta}_y) p(\vec{\theta}_y)}{P(R_x = k)} \quad (6)$$

$$= \frac{\mathbf{E}\{\theta_{xk} \mid \vec{\theta}_y\} p(\vec{\theta}_y)}{\mathbf{E}\{\theta_{xk}\}}$$

To clarify the functional relationships it is convenient to define

$$\mathrm{M}_k(\vec{\theta}_y; x, y) := \begin{cases} \theta_{xk} & , x = y \\ \mathrm{E}\{\theta_{xk} \mid \vec{\theta}_y\} & , x \neq y \end{cases} \quad (7)$$

Then, the posterior density $\vec{\theta}_y$ given the definitive observation is simply expressed in terms of $\mathrm{M}_k(\vec{\theta}_y; x, y)$

$$p(\vec{\theta}_y \mid R_x = k) = \frac{\mathrm{M}_k(\vec{\theta}_y; x, y) p(\vec{\theta}_y)}{\mathbf{E}\{\theta_{xk}\}} \quad (8)$$

The function $\mathrm{M}_k(\vec{\theta}_y; x, y)$ is fundamental to determine how a rock find at the location x affects the rock ratio's at all other locations y (including x). In recognition of the functions importance it will be henceforth referred to as a *geographical model*, as it describes the statistical relationships between samples at different geographic positions.

Properties of $\mathrm{M}_k(\vec{\theta}_y; x, y)$

The geographical model $\mathrm{M}_k(\vec{\theta}_y; x, y)$ is a complicated multi-variate function, determined by the underlying geological (or other) mechanism by which different rock types are distributed across an area. Nonetheless, applying reasonable assumptions it is possible to constrain it and make it tractable without going into the details of the (unknown) underlying mechanism.

- *Small influence at a distance*

Observations at a distant location should have little or no relation to the rock ratios at the current location. This assumption can be formalized as

$$\mathrm{M}_k(\vec{\theta}_y; x, y) \xrightarrow[|x-y| \to \infty]{} \mathrm{E}\{\theta_{xk}\} \quad (9)$$

Following this it also reasonable to assume some finite cut-off distance, beyond which measurements have no effect:

For some distance D > 0 (10)

$$|x - y| > D \Rightarrow \mathrm{M}_k(\vec{\theta}_y; x, y) = \mathrm{E}\{\theta_{xk}\}.$$

- *Decreasing influence with distance*

Not only should observations at locations distant to each other be largely independent, but the dependence between observations should never increase when the distance between them is increased. While motivated by the previous statement, this is a stronger assumption. Formally:

For any locations x, y, z s.t. $|x-y| > |x-z|$ (11)

$$\left| \mathrm{M}_k(\vec{\theta}_y; x, y) - \mathrm{E}\{\theta_{xk}\} \right| \leq \left| \mathrm{M}_k(\vec{\theta}_z; x, z) - \mathrm{E}\{\theta_{xk}\} \right|.$$

- *Smoothness assumption*

$\mathrm{M}_k(\vec{\theta}_y; x, y)$ should vary smoothly with x and y. This is consistent with natural laws and not very restrictive. However, relaxing this slightly to allow a discontinuity at the finite cut-off distance may be convenient.

- *Spatial invariance and isotropy*

This assumption requires that $\mathrm{M}_k(\vec{\theta}_y; x, y)$ be solely a function of the distance $|x-y|$ between samples, and not depend on their individual positions or the direction from one to the other. However, this is inconsistent with the previous assumptions, which require explicit dependence on x. Nonetheless, if this dependence can be explicitly accounted for and functionaly separated from

the rest of $M_k(\vec{\theta}_y; x, y)$, then spatial invariance and isotropy are desirable and reasonable properties, provided that the geographic area of interest is not too large.

- *Conjugacy requirement*

This is the most restrictive and useful of the assumptions, requiring that for all positions x,*y*, and rock types k, the prior distribution of rock ratios $p(\vec{\theta}_y)$ is the same class of distribution as the posterior $p(\vec{\theta}_y \mid R_x = k)$. Conjugate prior distributions are computationally convenient and the prior is easily interpretable as earlier measurements.

It is natural to represent $p(\vec{\theta}_y)$ by a Dirichlet distribution [7] with parameters $\alpha_{y1} \ldots \alpha_{yM}$:

$$p(\vec{\theta}_y) = \text{Dirichlet}(\vec{\theta}_y; \alpha_{y1} \ldots \alpha_{yM}) \tag{12}$$

$$= \frac{\Gamma(\alpha_{y1} + \ldots + \alpha_{yM})}{\Gamma(\alpha_{y1}) \ldots \Gamma(\alpha_{yM})} \theta_{y1}^{\alpha_{y1}-1} \ldots \theta_{yM}^{\alpha_{yM}-1}$$

Ensuring that $p(\vec{\theta}_y \mid R_x = k)$ is also a Dirichlet distribution requires that

$$M_k(\vec{\theta}_y; x, y) = \tag{13}$$

$$Z_k(x, y) \theta_{y1}^{\beta_1(x,y,k)} \ldots \theta_{yM}^{\beta_M(x,y,k)}$$

which guarantees from (8) that

$$p(\vec{\theta}_y \mid R_x = k) = \tag{14}$$

$$\text{Dirichlet}(\vec{\theta}_y; \alpha_{y1} + \beta_1(x,y,k), \ldots, \alpha_{yM} + \beta_M(x,y,k))$$

The small influence at a distance assumption (9) implies that

$$Z_k(x, y) \xrightarrow[|x-y| \to \infty]{} E\{\theta_{xk}\} \tag{15}$$

$$\beta_i(x, y, k) \xrightarrow[|x-y| \to \infty]{} 0$$

Introducing the finite cutoff of (10) implies the limits are attained when x and y are separated by a finite distance. The assumption of decreasing influence with distance (11) means that Z_k and $\beta_1 \ldots \beta_M$ never get further from their limits as $|x-y|$ increases. They are smooth functions as $M_k(.; x, y)$ is assumed smooth.

Note that computation of the posterior distribution requires only that $\beta_1(x,y,k) \ldots \beta_M(x,y,k)$ be specified. $Z_k(x,y)$ is implicitly defined by normalization of the posterior in (8). Furthermore, $Z_k(x,y)$ accounts for the spatial dependence of $M_k(.; x, y)$, making it possible to assume spatial invariance and isotropy for $\beta_1(x,y,k) \ldots \beta_M(x,y,k)$. That is, $\beta_j(x, y, k) = \beta_j(|x-y|, k)$.

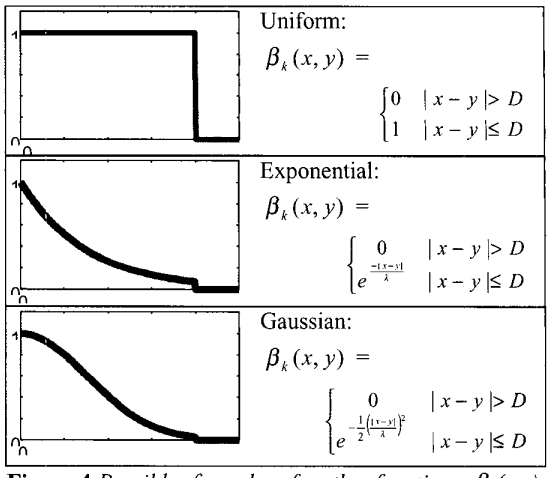

Figure 4 *Possible formulae for the functions $\beta_k(x,y)$. While ad hoc, they satisfy all the restrictions on $\beta_k(x,y)$, including spatial invariance and isotropy. The uniform formula weights all samples within an area equally, whilst the others have steadily decreasing influence with distance. Choice of the constants D and λ involves a trade-off between spatial resolution and power to generalize. Large values allow the rapid learning of the rock ratios over a large area from a few samples but are less effective in learning regional variations, which will be smoothed over. Ideally, they should be comparable to the average separation between samples.*

For co-located samples $x = y$, hence $M_k(\vec{\theta}_y; x, y) = \theta_{yk}$, implying that

$$\beta_k(x,x,k) = 1 \text{ and } \beta_j(x,x,k) = 0 \text{ (for every } j \neq k) \tag{16}$$

Because $\beta_1(x,y,k) \ldots \beta_M(x,y,k)$ approach zero and may not increase as distance $|x - y|$ increases, it follows that $\beta_j(x,y,k) = 0$ for all x, y and $j \neq k$. Therefore the notation for the β's is redundant; it is sufficient to denote the remaining nonzero term $\beta_k(x,y,k)$ as $\beta_k(x,y)$.

This implies that a *definitive* rock find (identified with 100% certainty) will not increase the assumed probabilities of finding other related but distinct rock types in the area.

To summarize

$$M_k(\vec{\theta}_y; x, y) = Z_k(x, y) \theta_{yk}^{\beta_1(x,y)} \tag{17}$$

and

$$p(\vec{\theta}_y \mid R_x = k) = \quad (18)$$

Dirichlet($\vec{\theta}_y$; α_{y1} ,..., $\alpha_{yk} + \beta_k(x,y)$, ..., α_{yM})

To learn the distribution of rock ratios at a particular point given definitive observations in the area, sum the contributions of each observation using (18) and an appropriate formula for the β's (Figure 4).

Learning from uncertain observations

Regrettably, it is rarely the case that rock samples can be autonomously identified with certainty. Otherwise, there would be little reason to learn the priors.

Consider an observation O_x made on a rock sample R_x at position x, with likelihoods w_{xk}=P(O_x|R_x=k). It can be shown that

$$p(\vec{\theta}_y \mid O_x) \propto \sum_k [P_{xk} Z_k(x,y) \theta_{xk}^{\beta_k(x,y)}] p(\vec{\theta}_y) \quad (19)$$

where

$$P_{xk} = P(R_x = k \mid O_x) \quad (20)$$

$$= \frac{P(O_x \mid R_x = k)\, \mathrm{E}\{\theta_{xk}\}}{\sum_j P(O_x \mid R_x = j)\, \mathrm{E}\{\theta_{xj}\}} \text{ (Bayes rule)}$$

$$= \frac{w_{xk}\alpha_{xk}}{\sum_j w_{xj}\alpha_{xj}}$$

As $p(\vec{\theta}_y)$ is a Dirichlet distribution, it follows that

$$p(\vec{\theta}_y \mid O_x) = \quad (21)$$

$$\sum_{k=1}^{M} P_{xk} \text{ Dirichlet}(\vec{\theta}_y ; \alpha_{y1},..,\alpha_{yk}+\beta_k(x,y),.., \alpha_{yM})$$

This is not a Dirichlet distribution, but a mixture model of Dirichlet distributions. It violates the conjugacy requirement and is intractable as more observations are made. Subsequent observations would produce mixture models with M^2, M^3, and so on terms. To maintain a closed form model with a bounded number of parameters it is necessary to approximate

$$P(\vec{\theta}_y \mid O_x) \quad (22)$$

Dirichlet($\vec{\theta}_y$; $\alpha_{y1}+P_{x1}\beta_1(x,y)$,..,$\alpha_{yM}+P_{x1}\beta_M(x,y)$)

This is equivalent to computing the posterior distribution of $\vec{\theta}_y$ given the definite observations $\{\mathbf{R_x}=k|k=1...M\}$ at x, each weighted according to their probability $P(\mathbf{R_x}=k|\mathbf{O_x})$ given the current observation and priors on $\vec{\theta}_y$. This exactly equals the true posterior whenever the observations identify the rock with complete certainty. The approximation is worst when the observations do not favor any rock type. The latter occurs when an observation is made that is equally likely for all possible rock types, in which case the probability of each rock type is equal to the prior. The approximation (21) of the posterior distribution on $\vec{\theta}_y$ then correctly has the same mean as the prior but a slightly decreased variance, which is incorrect as nothing has been learned.

Multiple rock finds

Assuming the prior rock ratio density at any point x is Dirichlet, with parameters α_{xk}^{prior}, and an observation is made on a particular sample. Then, the rock type probabilities for that sample are given by (20), and the rock ratio densities at all other points given by (22). If observations are made at all sample locations then (20) and (22) are combined to create the following coupled simultaneous equations, relating the rock type probabilities at each sample and the posterior rock ratio distribution at each sample given the rock type probabilities at the other samples:

$$P_{xk} = \frac{w_{xk}\alpha_{xk}}{\sum_j w_{xj}\alpha_{xj}}, \quad \alpha_{xk} = \alpha_{xk}^{prior} + \sum_{y \neq x} P_{yk}\beta_k(y,x) \quad (23)$$

The α_{xk}^{prior}'s define the assumed (Dirichlet) distributions on $\vec{\theta}_x$ at all sample locations x prior to any measurements.

Note that the order in which samples are examined does not affect the computed probabilities and rock ratio distributions.

Experimental results

Results obtained by applying (23), using Gaussian geographical models (Figure 4) to simulated data have shown statistically robust improvements in classification accuracy and generated consistent probability maps. However, there are too many arbitrary parameters to set when simulating data for it necessarily to be a good indicator of performance. There is no substitute for *real* data, gathered in the field by a robot.

In December 1998 the Nomad robot was dispatched to Patriot Hills, Antarctica where it traversed a moraine and acquired spectral data, using a fiber optic reflectance spectrometer operating in the visible range, from 51 rocks in its path, along with their spatial coordinates (Figure 5). These rock samples were recovered and

subsequently classified by a geologist to ascertain ground truth.

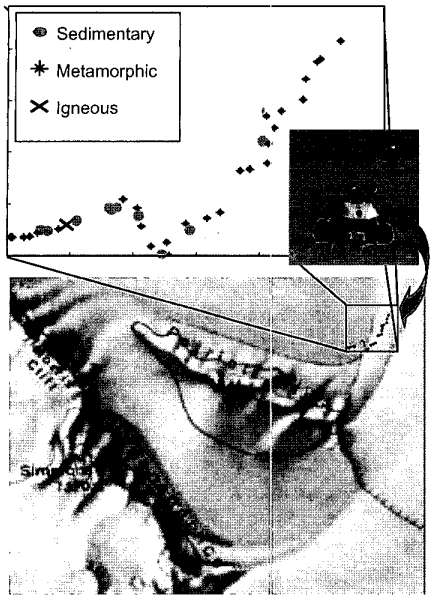

Figure 5 *Patriot Hills, Antarctica, data collection site. The Nomad robot traversed a 600m x 300m area, collecting spectroscopic data from each rock along it's path. The samples have been grouped into according to their formation process: sedimentary, metamorphic, igneous and extraterrestrial (meteorites, although none are present in the sample), each of which encompasses many non-overlapping rock types.*

Each sample spectrum was processed and run through a Bayes network classifier developed for rock and meteorite identification [3] to determine the conditional likelihood of that spectrum for each of 25 different rock types. This data was sequentially entered into a statistical model (22) of the possible rock ratio's at each sample position, using a Gaussian geographical model (Figure 4) and assuming initial distributions with α_{xk}^{prior}'s = 0.1 at all locations. At each iteration the rock type probabilities for every sample entered so far were recomputed and the most likely formation (sedimentary, metamorphic, igneous and extraterrestrial) deduced. Comparing with the known formation processes, the cumulative number of classification errors Figure 6(ii) as each sample is entered are determined.

Note the occasional *reduction* in the total number of misclassifications as more samples are examined and the model becomes more precise. This would not be possible if each rock sample was examined independently, as in Figure 6(i) and Figure 6(v) where the rocks are respectively classified with the assumption of uniform fixed rock type priors everywhere, and the priors fixed to the known data set rock type ratios. These indicate the worst and the best that the rock classifier in [3] can do under different assumptions on the rock type priors over the entire region if each rock is classified independently and the priors are assumed the same at each sample location.

Note that classification performance improves for various values of λ, indicating robustness to the exact form of the geographic models. Further computations using a uniform geographic model empirically confirm this. Furthermore, for large values of λ, Bayes optimal classification performance is *exceeded*. While from this data set it is hard to tell if the improvement is statistically significant, it is not inconsistent as the system is exploiting dependencies between samples as well as learning their geographical distributions.

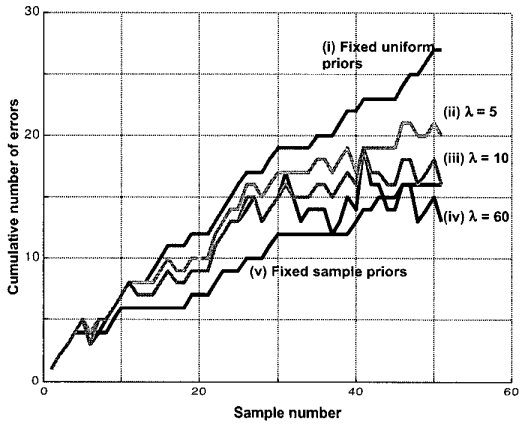

Figure 6 *Cumulative number of misclassifications as samples from the Antarctic field data are examined in the order they were encountered by the robot. In (i) samples are each independently classified assuming uniform rock type probabilities. In (ii)-(iv), the rock type probabilities are learned as samples are acquired, using Gaussian geographic models with λ values of 5, 10 and 60 respectively (c.f. Figure 4). Curve (v) indicates performance when rocks are again classified independently, using the known fraction of each rock type in the data set as the priors. This is the best performance possible for the independent classification of the rocks.*

With all the rock data entered into the model it is also possible to compute the learned rock type probabilities everywhere in an area (Figure 7, Figure 8, and Figure 9),

not just at the samples. Compare these to the actual distribution of the rock samples along the robots path in Figure 5 to verify that they are indeed consistent with what was seen by the robot. The regions dominated by sedimentary and metamorphic rocks are clearly identified. Conversely, nothing is learned about the areas distant from the robot path, since a geographic model with a finite cut-off is assumed. Furthermore, the low density of igneous rocks (and meteorites, which have a probability map almost identical to Figure 9) is learned, even in the area dominated by metamorphic rocks, which are often confused with igneous rocks.

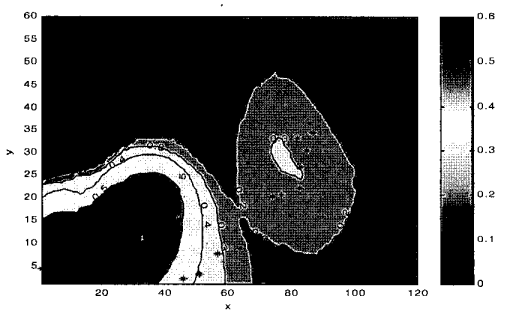

Figure 7 *Learned probability of sedimentary rocks across the explored region of the data collection site in Antarctica. The original sample positions and types are indicated by the colored dots (c.f. Figure 5).*

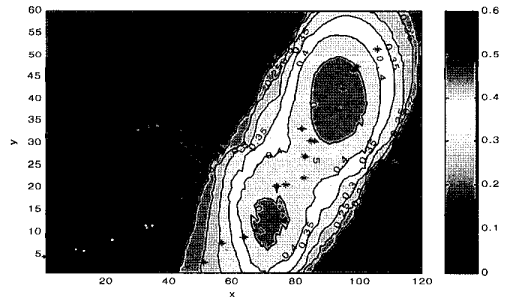

Figure 8 *Learned probability of metamorphic rocks.*

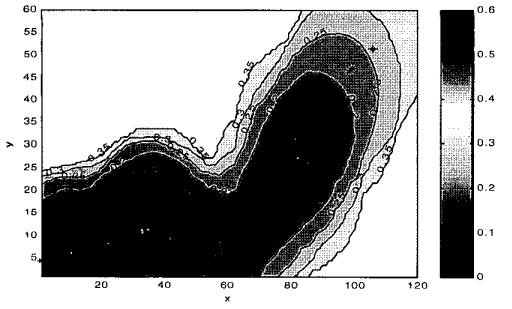

Figure 9 *Learned probabilities of igneous rocks.*

Conclusions

The results of Figure 6 are very significant. They show that learning the probabilities gives a clear improvement in classification over assuming uniform priors everywhere and classifying each sample independently. In fact, performance approaches and possibly exceeds the optimum, achieved when the average priors over the region are known beforehand. The latter might occur with more data and pronounced regional variations that can be exploited by this approach.

This is equivalent to a human geologist who looks at many rocks, constantly re-evaluating all the previous rocks seen every time another rock is looked at. Unfortunately, because of the lack of prior work in this exact area it is not possible to compare this with other methods.

A weakness of the approach proposed here is the ad hoc nature of the geographical models. While based upon reasonable assumptions they are still under-constrained. Nonetheless, empirical evidence suggests that the improvements in classification are robust to changes in the assumed geographical models. Choosing models with a wide footprint (large cut-off distance and low decay rate) results in faster convergence and generalizes over a larger area, but also less ability to capture and exploit small scale variations. Further work is needed to determine the optimal geographic models from the data.

All rock type probabilities are re-computed every time another rock sample is examined, and do not depend on the order in which they are found. Therefore, this method is robust to unlucky sequences of samples not representative of the area. Except for the approximation (22), the learning algorithm is Bayesian, and should converge (in the probabilistic sense) to the correct probabilities as more data is added.

Computationally requirements are minimal. They increase with the number of samples squared. In practice, the matrices in (23) are sparse and, depending on the robot path, complexity is order N^2 each time a new sample is added, where N is the average number of samples within a circle whose radius is the geographical model cut-off distance.

Acknowledgements

Invaluable assistance was provided by William Cassidy, Ted Roush, Andrew Moore, Martial Hebert, Dimitrious Apostolopoulos, Peter Cheeseman and all the participants in CMU's RAMS program. This work was supported by a grant from NASA.

References

[1] D. Apostolopoulos, M. Wagner, B. Shamah, L. Pedersen, K. Shillcutt, W. Whittaker, "Technology and Field Demonstration of Robotic Search for Antarctic Meteorites", *International Journal of Robotic Research*, special Field & Service Robotics issue, in press December 2000.

[2] L. Pedersen, D. Apostolopoulos, W. Whittaker, T. Roush and G. Benedix "Sensing and data classification for robotic meteorite search" Proceedings of *SPIE Photonics East Conference*. Boston, USA, 1998.

[3] L. Pedersen, D. Apostolopoulos, W. Whittaker "Bayes Networks on Ice: Robotic Search for Antarctic Meteorites", *Neural Information Processing Symposium 2000*, Denver, Colorado, 2000

[4] A. Elfes, "Occupancy rids: a stochastic spatial representation for active robot perception", *Autonomous Mobile Robots: Perception, Mapping and Navigation,* S. Iyengar, A. Elfes, Eds. IEEE Comp. Soc. Press, 1991.

[5] E. Gelenbe, N. Schmajuk, J. Staddon, J. Reif, "Autonomous search by robots and animals: a survey", *Robotics and Autonomous Systems* 22, 1997, pp 23-34.

[6] E. Gelenbe, Y. Cao, "Autonomous search for mines", *European Journal of Operational Research* 108, 1998, pp319-333.

[7] A. Gelman, J. Carlin, H. Stern, D. Rubin, *Bayesian Data Analysis,* Chapman & Hall, 1995 edition.

[8] Pearl, J., "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference", Morgan Kaufman, 1988.