ESTIMATING BEDROCK AND SURFACE LAYER BOUNDARIES AND CONFIDENCE INTERVALS IN ICE SHEET RADAR IMAGERY USING MCMC

Stefan Lee   *Jerome Mitchell*   *David J. Crandall*   *Geoffrey C. Fox*

School of Informatics and Computing
Indiana University
Bloomington, Indiana USA

ABSTRACT

Climate models that predict polar ice sheet behavior require accurate measurements of the bedrock-ice and ice-air boundaries from ground-penetrating radar imagery. Identifying these features is typically performed by hand, which can be tedious and error prone. We propose an approach for automatically estimating layer boundaries by viewing this task as a probabilistic inference problem. Our solution uses Markov-Chain Monte Carlo to sample from the joint distribution over all possible layers conditioned on an image. Layer boundaries can then be estimated from the expectation over this distribution, and confidence intervals can be estimated from the variance of the samples. We evaluate the method on 560 echograms collected in Antarctica, and compare to a state-of-the-art technique with respect to hand-labeled images. These experiments show an approximately 50% reduction in error for tracing both bedrock and surface layers.

Index Terms— Polar Science, Radar Imagery, Bedrock and Surface Layers, Probabilistic Graphical Models

1. INTRODUCTION

Observing the structure and dynamics of the polar ice sheets is critical for developing accurate climate models. Traditionally, glaciologists have used ice core drilling practices to understand the subterranean structure of an ice sheet, but fortunately, ground-penetrating radars have matured to allow surveying large areas of ice from aerial and ground vehicles. For instance, the Center for Remote Sensing of Ice Sheets (CReSIS) (among other groups) has developed and deployed a multichannel coherent radar depth sounder to image cross sections of the ice. Figure 1 shows an example of a radar echogram produced by this process, where the horizontal axis is distance along a flight line and the vertical axis is vertical distance (depth) from the plane. Echograms can be used to estimate the topography of the bedrock below the ice (which is the dark erratic line near the middle of the figure), and the depth of the ice (i.e. the distance from the bedrock to the line near the top of the echogram, which is the ice surface). These
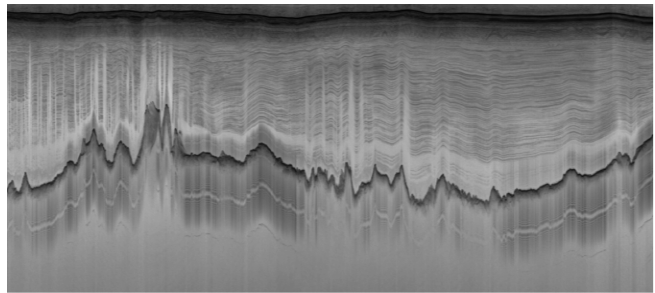

Fig. 1. Sample radar echogram of an ice sheet, including the surface (very dark line near the top) and bedrock (dark erratic line near the middle) layer boundaries, along with weaker returns from contiguous layers of ice between the two.

observations can be used as input into glaciological models of ice sheet behavior over time.

Ground-penetrating radar has allowed for data to be collected across vast areas of ice, but analyzing it remains a challenge, and is typically done by hand. A few recent papers have studied how to use image processing and computer vision techniques to find layer boundaries (semi-)automatically from echograms [1, 2, 3, 4, 5, 6, 7, 8, 9], but this is a hard problem because of the high degree of noise, the often faint layer boundaries, and confusing linear structures that are caused by signal reflections and clutter. In fact, even human labelers produce diverging estimates of the boundaries in many cases. We thus need new techniques that combine together weak image cues, reasoning explicitly about uncertainty in both the evidence and the resulting layer boundary estimates.

In this paper, we pose layer identification as an inference problem on a statistical graphical model, building on the approach introduced by Crandall et al [10]. The probabilistic framework allows for multiple sources of evidence to be integrated in determining layer boundary estimates (as opposed to, for instance, using an edge detection step, which determines hard and irreversible decisions at an early processing stage). We introduce several important contributions to improve both the accuracy and utility of layer-finding. Our technical innovation uses Gibbs sampling for performing

inference instead of the dynamic programming-based solver of [10]. This allows us to strengthen the underlying model and solve for layer boundaries simultaneously, yielding automatic layer detection results that are significantly better than the approach in that paper. Moreover, the Gibbs sampler produces explicit confidence intervals, thus giving bands of uncertainty in the layer boundary locations. Since noise and ambiguity in radar echograms are inevitable, we believe this ability to determine confidences could be crucial in applications of layer identification (e.g. when used in glaciological models), and to our knowledge this is the first paper that has demonstrated this concept on the layer-finding problem.

2. RELATED LITERATURE

Several semi-automated and automated methods for identifying subsurface features of ice have been introduced in the literature. The most related papers to our work have focused on automated detection in terrestrial echograms. For example, Freeman et al. [5] and Ferro and Bruzzone [3] investigated how shallow ice features can be automatically detected in icy regions on Mars. In other work, Ferro and Bruzzone used echograms of the Martian subsurface to detect basal returns. The subglacial identification problem was studied by Gifford et al [1], who compared two primary approaches, namely an active contour ("snake") model and an edge based technique. Ilisei et al. [4] developed a two phase technique to exploit the properties of a radar signal for generating a statistical map and applying a segmentation algorithm. Although our application focuses on detecting bedrock and surface layers, other studies use similar techniques to identify internal layers in radar imagery. Approaches include Fahnestock et al. [7], Karlsson and Dahl-Jensen [6], Sime et al. [8], and Mitchell et al. [11].

Our approach is most closely related to Crandall et al. [10], and we use a similar probabilistic framework here. However, our model makes fewer assumptions and our inference algorithm is able to jointly solve for all layer boundaries simultaneously, and our experiments show a significant improvement in quantitative accuracy compared to ground truth. Additionally, our approach is able to characterize uncertainty of the layer boundary estimates by calculating confidence intervals, whereas the technique in [10] simply gives a single layer boundary with no measure of certainty.

3. METHODOLOGY

An echogram is a 2D matrix which represents the scattering properties of the subsurface at each along-track coordinate of the platform. Figure 1 shows an example of an echogram from the CReSIS dataset [12]. Our task is to find two key features in these echograms: the ice surface boundary (the strong reflector near the top) and the bedrock boundary (the dark reflector near the middle of the image).

3.1. Modeling layer boundaries

We want to estimate the location of layer boundaries by determining their paths through the image. Assume that an echogram has k layer boundaries (k=2 in our case). Given an echogram I of dimension $m \times n$, we wish to estimate latent variables $L_i = \{l_{i1}, ..., l_{in}\}$, where l_{ij} denotes the row coordinate of layer $i \in [1, k]$ in column j of the image.

We take advantage of the structure of this problem by posing it as a grid-shaped probabilistic graphical model. In this framework, we are interested in estimating $P(L_1, ..., L_k | I)$, the joint probability over the layer boundaries given the echogram. Unfortunately, this distribution has an alarming dimension of order $O(m^{Kn})$ so that computation and storage is intractable even for small images. To address this problem, we make three simplifying assumptions: (1) all echograms are equally likely; (2) image characteristics are determined by local layer boundaries; and (3) variables in L exhibit a Markov property with respect to their local neighbors.

Under the first assumption, the joint distribution can be factored into a product according to Bayes' Law,

$$P(L_1, ..., L_k | I) = P(I | L_1, ..., L_k) P(L_1, ..., L_k). \quad (1)$$

This decomposition reduces the full joint into two intuitive distributions: $P(I | L_1, ..., L_k)$ captures how well the image data can be explained by a set of layers $L_1, ..., L_k$, and $P(L_1, ..., L_k)$ captures prior knowledge about the boundaries, like that they are smooth and do not intersect.

The second assumption implies that parts of the image not near the layer boundaries are generated by noise, so we need only model pixels near boundaries. Thus we can factor $P(I | L_1, ..., L_k)$ into a product over layers and columns,

$$P(I | L_1, ... L_l) = \prod_{i=1}^{k} \prod_{j=1}^{n} P(I | l_{i,j}). \quad (2)$$

Since boundaries are dark edges, we model the right hand term as a product of gradient magnitude and image intensity,

$$P(I | l_{i,j}) \propto |\nabla I(l_{i,j}, j)| \cdot (1 - I(l_{i,j}, j)), \quad (3)$$

where $|\nabla I(x, y)|$ is the gradient magnitude at coordinate (x, y) of the image, and we assume that pixel values have been scaled such that $I(x, y) \in [0, 1]$. We approximate gradient magnitude using finite differences on a 5×5 window.

The third assumption simplifies the problem by assuming the graphical model has the property that each node $l_{i,j}$ is independent of the remaining variables in L given its immediate neighbors in the graph. Under this assumption, we have,

$$P(L_1, ..., L_k) \propto \prod_{i=1}^{k} \prod_{j=1}^{n} P(l_{i,j} | N(l_{i,j})) \quad (4)$$

where $N(l_{i,j})$ is the set of directly connected nodes in the graph (i.e. $N(l_{i,j}) = \{l_{a,b} \mid 1 = |a - i| \text{ and } 1 = |b - j|\}$). We

define $P(l_{i,j}|N(l_{i,j}))$ as the product of independent vertical and horizontal components. Along the same layer, l_i's are encouraged to be smooth by a zero-mean Gaussian which is truncated to zero outside a fixed interval,

$$P(l_{i,j}|l_{i,j-1}) = \begin{cases} \mathcal{N}(l_{i,j} - l_{i,j-1}; 0, \sigma) & |l_{i,j} - l_{i,j-1}| < \phi_H \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

while a step function encourages layers not to overlap,

$$P(l_{i,j}|l_{i-1,j}) \propto \begin{cases} 0 & l_{i,j} \le l_{i,j-1} \\ 0.1 & |l_{i,j} - l_{i,j-1}| < \phi_V \\ 1 & \text{otherwise.} \end{cases} \quad (6)$$

This model is similar to [10] but with important improvements. In [10], the vertical pairwise potentials are zero at and above intersection points and uniform elsewhere. But it is common in this data to see radar reflections of the surface layer directly below the actual surface, so we add a fixed-width low probability region directly below them to reduce false bedrock detections on these reflections. Perhaps more importantly, the model in [10] breaks these vertical constraints in order to simplify inference by greedily solving each layer conditioned on the previous one. We avoid doing this, and our experiments show that this holistic inference approach offers substantial improvement in accuracy.

3.2. Statistical inference

The model defined by equations (1), (2), and (4) is a first-order Markov Random Field. Unfortunately, finding the values of L that maximizes equation (1) is NP-hard in the general case [13]. Rather than trying to solve this as an optimization problem, we instead attempt to estimate functionals of the full joint distribution via Gibbs sampling.

Gibbs sampling is a Markov Chain Monte Carlo (MCMC) method which is capable of producing samples $X^{(1)}, ..., X^{(J)}$ from a distribution $f(x)$ without requiring the ability to directly sample or even know the form of $f(x)$ [14]. This is accomplished by iteratively sampling each variable conditioned on the remaining variables. Pseudo-code for Gibbs sampling is shown in Figure 2. This sampler provides a flexible framework for generating samples of a complex distribution, assuming samples can be taken from usually simpler full conditionals. The run-time complexity is $O(mJ)$, but in practice depends on the ease of sampling from the full conditionals.

It can be shown via Bayes Law and the independence assumptions in equations (2) and (4) that the full conditionals for each l_{ij} can be computed easily,

$$P(l_{ij}|I, N(l_{ij})) = P(I|l_{ij})P(l_{ij}|N(l_{ij})). \quad (7)$$

As the domain of l_{ij} is discrete and finite, sampling from this conditional is well-defined. As an additional optimization, we make use of the vertical and horizontal thresholds in equation (6) to sparsify the computation of $P(l_{ij}|N(l_{ij}))$, since

```
1: Initialize X^(0) = {x_1, ..., x_m};
2: j = 1;
3: while j < J do
4:     X^(j) = X^(j-1);
5:     for all x_i in X^(j) do
6:         x_i^(j) ~ P(x_i|X^(j) - {x_i^(j)});
7:     end for
8:     j = j + 1;
9: end while
```

Fig. 2. General algorithm for Gibbs sampling.

most entries are known to be zero. We apply the Gibbs sampler to generate a sequence of samples $L^{(B)}, ..., L^{(J)}$ where B is a burn-in time during which samples are discarded. This is a common practice with MCMC methods to reduce sensitivity to initial values.

To predict the layer locations we take the mean of our $M = J - B$ samples, which approximates the expectation of the joint distribution for large M,

$$E\left[P(L_1, ..., L_k|I)\right] = \lim_{M \to \infty} \frac{1}{M} \sum L^{(i)}.$$

Another helpful property of sampling is that the marginal distributions can be derived simply by discarding the other variables in the sample. We utilize this to produce point-wise 95% confidence intervals by taking the 0.025 and 0.975 quantiles of these approximate marginal distributions.

4. EXPERIMENTAL RESULTS

We tested our layer-finding approach using a set of 826 publicly-available radar echograms from the 2009 NASA Operation Ice Bridge program, collected with the airborne Multichannel Coherent Radar Depth Sounder system described in [12]. Each echogram has a resolution of 700 by 900 pixels (where 900 pixels represents about 30km of data on the horizontal axis, and 700 pixels corresponds to 0 to 4 km of ice thickness on the vertical axis). We chose this dataset because it was used by [10], so that we can directly compare the accuracy of the two techniques, and we used the source code provided by the authors. (We plan to make the source code of our new technique publicly available as well.)

The images have ground-truth labels that were produced by human annotators, but these labels are often quite noisy. For instance, sometimes the annotators could not find a reasonable layer boundary and simply "gave up" by not marking anything at all. To decouple the error in the ground-truth from the method evaluation, we removed such images with incomplete ground truth (including those with partially defined layers and those in which less than two layer boundaries were specified). We ran our method on the remaining 560 images. For each image, we collected 10,000 samples after a burn-in period of B=20,000 iterations.
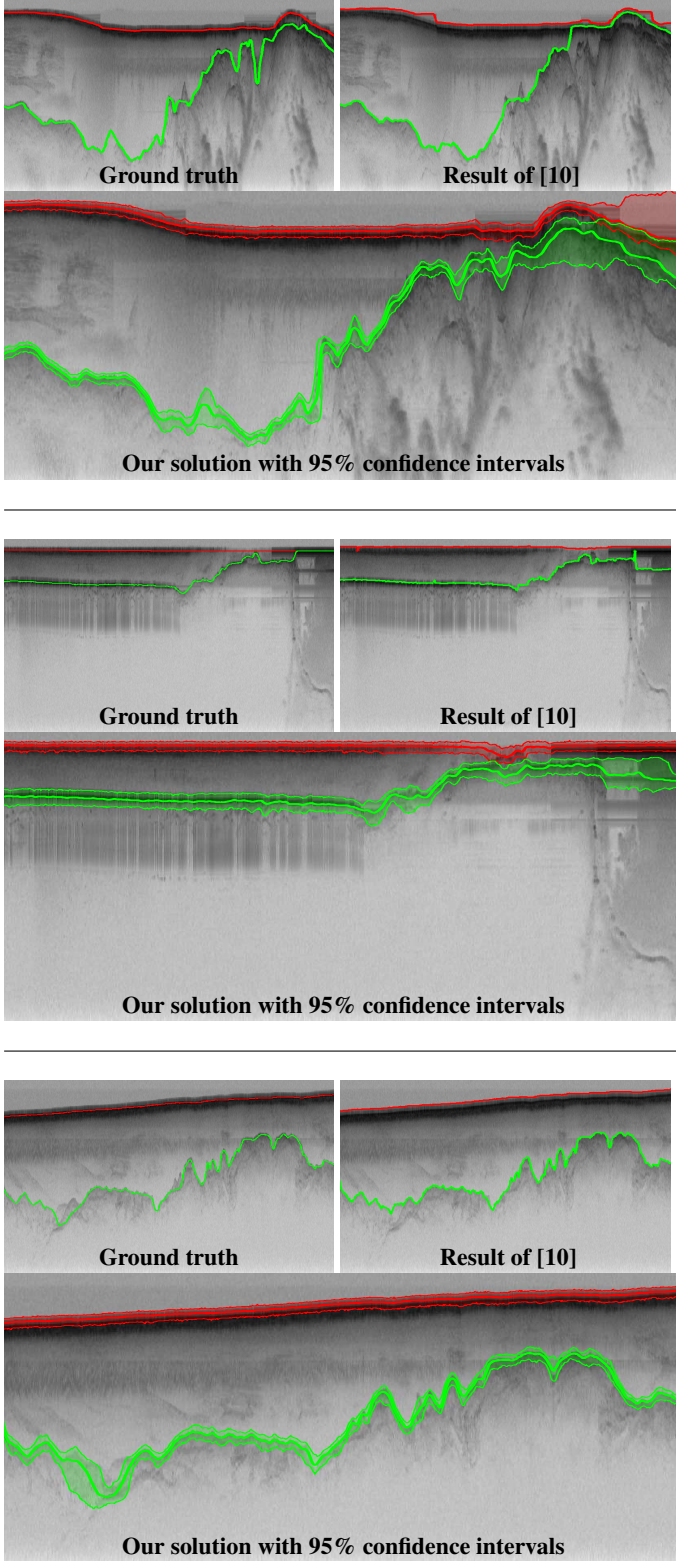
Fig. 3. Results on three sample echograms. Each pane includes the hand-labeled ground truth image *(top-left)*, the output of [10] *(top-right)*, and then our output *(bottom)*. **Best viewed in color.**

Approach	Mean Error		Median Mean Error	
	Surface	Bedrock	Surface	Bedrock
[10]	22.3	43.1	10.6	14.4
Ours	9.3	37.4	5.9	9.1

Table 1. Evaluation of our method on the test set. Error is measured in terms of absolute columnwise difference compared to ground truth, summarized with average mean deviation and median mean deviation across images, in pixels.

Figure 3 shows results on three sample echograms, presenting the output of our technique (including the confidence interval) as well as the ground truth and baseline technique of [10]. We present quantitative performance metrics in Table 1. We measure accuracy by viewing ground truth and estimated layer boundaries as 1-D signals, and computing the mean absolute deviation (in pixels) between the two. We use two summary statistics: mean columnwise absolute error over all images and the median of the columnwise mean absolute errors across images. The first measures how well predicted layers match the ground truth, treating columns within an image as uncorrelated, while the later metric recognizes that high error in one column in an image is highly correlated with the error in the remaining columns and looks at error from a per-image viewpoint. Under both metrics, we outperform the method of [10] significantly, by decreasing the error rate by about 44.3% for surface boundaries and 48.3% for bedrock.

We also quantified how informative the confidence intervals are by computing the percentage of ground truth layer points that are contained within the estimated intervals. We found that 94.7% of the surface boundaries and 78.1% of the bedrock boundaries are within the intervals, for an overall percentage of 86.4%. The fact that this number is close to but less than 95% reflects that our framework is a good but not perfect model of layers in echogram images.

5. CONCLUSION

We proposed an automated approach to estimate bedrock and surface layers in multichannel coherent radar imagery and demonstrated its effectiveness on a real-world dataset against the state-of-the-art. our technique produces confidence interval estimates and we evaluated their correctness. We believe layer-finders that provide such confidences may help improve climate models by quantifying error in the input data.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] C. Gifford, G. Finyom, M. Jefferson, M. Reid, E. Akers, and A. Agah, "Automated polar ice thickness estimation from radar imagery," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2456–2469, 2010.

[2] A. Ferro and L. Bruzzone, "Analysis of radar sounder signals for the automatic detection and characterization of subsurface features," *IEEE Transactions on Geoscience and Remote Sensing*, 2012.

[3] A. Ferro and L. Bruzzone, "Automatic extraction and analysis of ice layering in radar sounder data," *IEEE Transactions on Geoscience and Remote Sensing*, 2013.

[4] A.-M. Ilisei, A. Ferro, and L. Bruzzone, "A technique for the automatic estimation of ice thickness and bedrock properties from radar sounder data acquired at Antarctica," in *IEEE International Geoscience and Remote Sensing Symposium*, 2012, pp. 4457–4460.

[5] G. Freeman, A. Bovik, and J. Holt, "Automated detection of near surface Martian ice layers in orbital radar data," in *IEEE Southwest Symposium on Image Analysis & Interpretation*, 2010, pp. 117–120.

[6] N. Karlsson, D. Dahl-Jensen, S. P. Gogineni, and J. D. Paden, "Tracing the depth of the holocene ice in north greenland from radio-echo sounding data," *Annals of Glaciology*, 2012.

[7] M. Fahnestock, W. Abdalati, S. Luo, and S. Gogineni, "Internal layer tracing and age-depth-accumulation relationships for the northern greenland ice sheet," *Journal of Geophysical Research*, vol. 106, no. D24, pp. 33789–33, 2001.

[8] L. Sime, R. Hindmarsh, and H. Corr, "Instruments and methods automated processing to derive dip angles of englacial radar reflectors in ice sheets," *Journal of Glaciology*, vol. 57, no. 202, pp. 260–266, 2011.

[9] J. Mitchell, D. Crandall, G. Fox, M. Rahnemoonfar, and J. Paden, "A semi-automatic approach for estimating bedrock and surface layers from multichannel coherent radar depth sounder imagery," in *SPIE Conference on Remote Sensing*, 2013.

[10] D. Crandall, G. Fox, and J. Paden, "Layer-finding in radar echograms using probabilistic graphical models," in *International Conference on Pattern Recognition*, 2012, pp. 1530–1533.

[11] J. Mitchell, D. Crandall, G. Fox, and J. Paden, "A semi-automatic approach for estimating near surface internal layers from snow radar imagery," in *IEEE International Geoscience and Remote Sensing Symposium*, 2013.

[12] C. Allen, L. Shi, R. Hale, C. Leuschen, J. Paden, B. Pazer, E. Arnold, W. Blake, F. Rodriguez-Morales, J. Ledford, et al., "Antarctic ice depthsounding radar instrumentation for the NASA DC-8," *IEEE Aerospace and Electronic Systems Magazine*, vol. 27, no. 3, pp. 4–20, 2012.

[13] Daphne Koller and Nir Friedman, *Probabilistic graphical models: principles and techniques*, MIT Press, 2009.

[14] George Casella and Edward I George, "Explaining the Gibbs sampler," *The American Statistician*, vol. 46, no. 3, pp. 167–174, 1992.