Conditional Random Fields for Rock Characterization using Drill Measurements

Sildomar T. Monteiro, Fabio Ramos and Peter Hatherly
Australian Centre for Field Robotics
School of Aerospace, Mechanical and Mechatronic Engineering
University of Sydney, NSW 2006, Australia
s.monteiro@acfr.usyd.edu.au

Abstract

Analysis of drill performance data provides a powerful method for estimating subsurface geology. While there have been studies relating such measurement-while-drilling (MWD) parameters to rock properties, none of them has attempted to model context, that is, to associate local measurements with measurements obtained in neighbouring regions. This paper proposes a novel approach to infer geology from drill measurements by incorporating spatial relationships through a Conditional Random Field (CRF) framework. A boosting algorithm is used as a local classifier mapping drill measurements to corresponding geological categories. The CRF then uses this local information in conjunction with neighbouring measurements to jointly reason about their categories. Model parameters are learned from training data by maximizing the pseudo-likelihood. The probability distribution of classified borehole sections is calculated using belief propagation. We present experimental results of applying the method to MWD data collected from a semi-autonomous drill rig at an iron ore mine in Western Australia.

1. Introduction

This paper presents work that is part of a larger project aimed at developing a fully autonomous, remotely operated mine. A main challenge of an autonomous mine is to build representations of the in-ground geology to determine the quantity and quality of the minerals of interest. The large semi-autonomous drill rigs used for drilling blast holes in mining are equipped with sensors which provide measurements while drilling (MWD) of drill performance and operation. These are used to control and monitor the drilling process. However, MWD data can also be used to relate drill performance to the physical properties of the rocks being drilled. The idea of relating drilling measurements to geotechnical properties of rocks has been considered before. Previous works focused on determining empirical indices for rock strength based on drilling parameters [1, 2]. Statistical methods and machine learning techniques have also been applied to relate MWD data and lithology [3–5]. However, previous methods do not model spatial dependencies of nearby geology. With an accurate assessment of lithology and rock strength, it is possible to determine optimal explosive load and distribution and to maximise the recovery of the desired rock types.

This paper aims to address the problem of jointly estimating the geology of neighbouring regions by exploiting spatial dependencies. A classification method that takes into account spatial relationships is proposed in the framework of Conditional Random Fields (CRFs). CRFs are undirected graphical models that are very powerful for modelling relational information; spatial relationships and other types of contextual information [6]. By directly modelling the conditional probability of the hidden states given the observations rather than the joint probability, CRFs avoid the difficult task of specifying a generative model for observations, as necessary in techniques such as Hidden Markov Models or Markov Random Fields. As a result, CRFs can handle arbitrary dependencies between observations, which give them substantial flexibility in modelling complex dependencies.

The proposed method applies the CRF framework in conjunction with boosting algorithms to provide classification. Boosting is a machine learning technique for supervised classification which has a sound theoretical foundation and provides probability estimates for each class [7]. It has become very popular due to many empirical studies showing that it tends to yield smaller classification error rates and be more robust to overfitting than competing methods such as Support Vector Machines or Neural Net-

This work was supported by the Rio Tinto Centre for Mine Automation and by the ARC Centres of Excellence programme funded by the Australian Research Council and the New South Wales State Government.

The authors would like to thank Pal Annette, James Batchelor and Charles McHugh, from Rio Tinto, for their advice and assistance with this work.

works [8]. Boosting is used in this work to non-linearly map drill measurements to the estimated geology. The set of boosting classified labels are used by the CRF framework to learn model parameters discriminatively. The resulting CRF model specifies the spatial relationship between MWD data.

2. Conditional Random Fields

CRFs were originally proposed for labelling relational data [9]. CRFs directly model $p(\mathbf{x}|\mathbf{z})$; the *conditional* distribution over the hidden variables $\mathbf{x}$ given observations $\mathbf{z}$, where $\mathbf{x} = \langle \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \rangle$, and $\mathbf{z} = \langle \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n \rangle$. The nodes $\mathbf{x}_i$, along with the connectivity structure represented by the undirected edges define a conditional distribution $p(\mathbf{x}|\mathbf{z})$ over the hidden states $\mathbf{x}$. The edges in the graph represent potential functions which map sensor measurements to non-negative numbers. By using log-linear combinations of potential functions where local potentials are denoted as $h(\mathbf{x}_i, \mathbf{z}_i)$ and pairwise potentials as $g(\mathbf{x}_i, \mathbf{x}_j)$, the conditional probability distribution is written as:

$$p(\mathbf{x} \mid \mathbf{z}) = \frac{1}{Z(\mathbf{z})} \exp \left\{ \sum_i \sum_{k=1}^{K_1} w_k^h h_k(\mathbf{z}_i, \mathbf{x}_i) + \sum_{i,j} \sum_{k=1}^{K_2} w_k^g g_k(\mathbf{x}_i, \mathbf{x}_j) \right\}, \quad (1)$$

where $\mathbf{w}^h$ is a vector with K_1 dimensions representing the weights for local potentials, $\mathbf{w}^g$ is a vector with K_2 dimensions representing the weights for the pairwise potentials and $Z(\mathbf{z})$ is a normalising partition function. The computation of the partition function can be exponential in the size of $\mathbf{x}$, therefore exact solutions might be unfeasible. However, for the particular graphical model used in this paper (linear chain), inference can be computed exactly in polynomial time.

2.1. Model Learning

The goal of CRF parameter learning is to determine the weights of the feature functions used in the conditional likelihood (1). CRFs learn these weights discriminatively by maximizing the conditional likelihood of labelled training data. While there is no closed-form solution for optimizing (1), it can be shown that (1) is convex relative to the weights $\mathbf{w} = \{\mathbf{w}^h, \mathbf{w}^g\}$. Thus, the global optimum of (1) can be found using a numerical gradient algorithm. Unfortunately, this optimisation runs an inference procedure at each iteration, which can be intractably inefficient in our case.

We therefore resort to maximizing the *pseudo-likelihood* of the training data, which is given by the sum of local likelihoods $p(\mathbf{x}_i \mid \mathrm{MB}(\mathbf{x}_i))$, where $\mathrm{MB}(\mathbf{x}_i)$ is the Markov blanket of variable $\mathbf{x}_i$: the set of the immediate neighbours of $\mathbf{x}_i$ in the CRF graph [10]. Optimisation of this pseudo-likelihood is performed by minimizing the negative of its log, resulting in the following objective function:

$$L(\mathbf{w}) = -\sum_{i=1}^{n} \log p(\mathbf{x}_i \mid \mathrm{MB}(\mathbf{x}_i), \mathbf{w}) + \frac{\mathbf{w}^T \mathbf{w}}{2\sigma^2}. \quad (2)$$

Here, the terms in the summation correspond to the negative pseudo log-likelihood and the right term represents a Gaussian shrinkage prior with variance σ^2. Without additional information, the prior mean is typically set to zero. In our approach, we use unconstrained L-BFGS [11], an efficient gradient descent method, to optimise (2). The key advantage of maximizing pseudo-likelihood rather than the likelihood (1) is that the gradient of (2) can be computed extremely efficiently, without running an inference algorithm. Learning by maximizing pseudo-likelihood has been shown to perform very well in different domains; see [12–14].

2.2. Inference

Once the parameters are learned, the model can be used to estimate the labels of new unlabelled data. This step is referred to as inference. Inference in CRFs can estimate either the marginal distribution of each hidden variable $\mathbf{x}_i$ or the most likely configuration of all hidden variables $\mathbf{x}$ (*i.e.*, MAP estimation), as defined in (1). Both tasks can be solved using *belief propagation* (BP), which works by sending local messages through the graph structure of the model [15]. Each node sends messages to its neighbours based on messages it receives and the clique potentials, which are defined via the observations and the neighbourhood relation in the CRF. BP generates exact results in graphs with no loops, such as chains or trees.

3. CRFs for rock classification

The CRF model can be applied to jointly reason about drilling measurements and neighbouring sections in the axial direction of a borehole by using a chain-like structure as illustrated in Fig. 1. The i^{th} section of a borehole composed of n sections is modelled as two interconnected nodes $\mathbf{z}_i$ and $\mathbf{x}_i$ representing, respectively, the set of drill measurements and the rock category. Drill measurements are considered observed variables and are represented by shadowed nodes. Borehole section categories, which are not observed, correspond to latent (hidden) variables and are represented by clear nodes. The relationships between nearby borehole sections are represented by edges connecting them.

The CRF model can employ arbitrary features functions to describe any particular property of the data. The next subsections describe the potential functions used in this study.

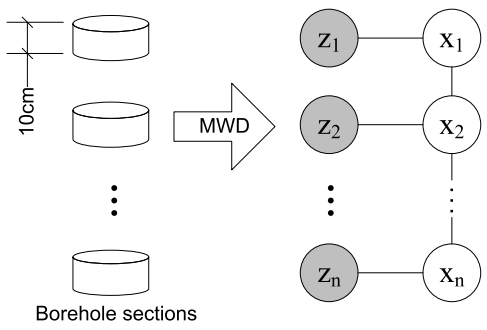


Figure 1. Graphical representation of the CRF framework modelling the spatial association between neighbouring borehole sections. The observations z_i correspond to drill measurements and the latent variables x_n indicate the corresponding classes.

3.1. Local features

In complex multimodal problems, instead of learning the CRF model directly from the raw observations (MWD data), it is advantageous to extract features from the data using a classification algorithm [14]. We used a boosting classifier to provide rock categories for each borehole section from the drill measurements. Boosting provides a nonlinear mapping from continuous variables to binary categories.

The most common version of boosting is AdaBoost [7]. The concept of AdaBoost is to train many weak learners on various distributions of the input data and then combine the classifiers produced into a single committee. We used a single node decision tree, also known as a decision stump, as weak learner [16]. Initially, the weights of all training examples are set equally, but after each round of the algorithm, the weights of incorrectly classified examples increase. The final committee or ensemble is a weighted majority combination of M weak classifiers and can be expressed as

$$h_k(\mathbf{z}) = sign\left(\sum_{m=1}^{M} \alpha_m^k C_m^k(\mathbf{z})\right), \qquad (3)$$

where α_m quantifies the contribution of each respective weak classifier C_m. In order to extend AdaBoost to multiclass problems, a one-against-all approach was implemented for each class k.

In addition, we implemented another boosting algorithm called LogitBoost, which fits additive logistic regression models by stagewise optimisation of the maximum likelihood [17]. LogitBoost can handle the multiclass problem directly.

Three ways of combining boosting and CRFs were investigated:

- The first uses the multiclass output of LogitBoost as continuous features in a CRF;
- In the second, the AdaBoost's weak learners for all classes are used as features. The weak learners' weight vector α is reset and the CRF learns corresponding weights instead;
- The third uses the AdaBoost's binary output h for each class k as features.

3.2. Pairwise features

A pairwise feature is used to associate measurements from neighbouring sections. The function associating a node $\mathbf{z}_i$ to a neighbouring node $\mathbf{z}_j$ is defined as

$$g(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} a & \text{if } \mathbf{x}_i = \mathbf{x}_j \\ b & \text{if } \mathbf{x}_i \neq \mathbf{x}_j \end{cases} \qquad (4)$$

where a and b are parameters learned from the data set.

4. Experimental Results

The effectiveness of the proposed method is evaluated using MWD data collected from an iron ore mine located in the Pilbara region of Western Australia. To provide training data for the algorithms, the drilled holes were tested using geophysical sensors: calliper, natural gamma, magnetic susceptibility and density (gamma gamma) logging tools. The detailed geology, as shown in Fig. 3(a) was determined by mine geologists using a combination of geological, geophysical, chip and core logs. Determining the detailed geology in terms of lithology, mineralogy and rock strength is a complex task and requires interpretation of the available data. This subjective process creates uncertainty in the training labels.

The algorithms were tested using k-fold cross-validation. In each round of training, instead of taking a random sample, all sections from one of the boreholes were left out for validation while a model was learned using the remaining data. The average of the classification performance over all validation holes was then calculated. This cross-validation approach was devised to simulate the real-world scenario of trying to estimate the geology of newly drilled boreholes using a model learned from previous holes.

The performance of the models was evaluated by calculating accuracy and F-measure. Accuracy, for an individual class, is the ratio of correct predictions in the population. F-measure is based on the harmonic mean of precision and recall. Precision indicates how many classified samples are correct and gives a measure of exactness. Recall, also called sensitivity, indicates how many of existing class samples the model returns and gives a measure of completeness. For multiple class problems, the overall performance can be calculated by two different types of average: *micro*-average

and *macro*-average. Micro-average gives equal weight to each sample whereas macro-average gives equal weight to each category. Micro-averaged accuracy is equivalent to the ratio of the number of correctly classified sections over the total number of samples.

The number of weak learners of the boosting algorithms can be determined experimentally. For comparison purposes, in all experiments the boosting algorithms were composed of 50 weak learners. Nevertheless, the boosting algorithm is quite resilient to overfitting and we observed that using more weak learners does not degrade performance severely.

4.1. Estimating detailed geology from percussion drilling

Drilling was conducted with a drill rig, shown in Fig. 2, operating in percussion mode and using a vibration shock absorber. In this experiment, a total of 12 drill measurements were recorded for analysis: bit air pressure, pull-down pressure, rotation pressure, pull-down rate, head speed, feed down pressure, feed up pressure, reverse rotation, forward rotation, rotation relief pressure, feed relief pressure, and hold back pressure. The data set consists of 28 boreholes drilled in a straight line 3 m apart and to a depth of 12 m. The boreholes' data are divided into 10 cm sections. The measurements, taken at different sampling frequencies, were re-sampled according to time stamp. Then, the average was calculated from measurements of each corresponding section. The geology of the target area can be categorised in four classes: banded iron formation (BIF), shale, and iron ore zones A and B. The corresponding label set is shown in Fig. 3(b).

Figure 4 shows an illustration of the geological sections for all boreholes comparing the reference labels used for model learning with the predictions of LogitBoost and the CRF methods. A quantitative analysis of the algorithms' performance is presented in Table 1, all the results are in percentages. All algorithms are able to learn a model from the available training area and to generalise the results to the whole data set. The CRF approach presents a "smoothing" effect which correlates better with the expected geology of the area. All methods seem to have some difficulty in learning the sloped boundary between different mineral zones, which might be a side effect of the way cross-validation was performed and nature of the training data set.

4.2. Estimating simple lithology from rotary and percussion drilling

For this experiment, 132 holes were drilled, 88 in rotary and 44 in percussion mode, and 5 drill measurements were recorded: bit air pressure, pull-down pressure, rotation pressure, pull-down rate, and rotation speed. In this site, detailed down-hole geophysical logs were not available, the geological interpretation was performed by mine geologists based on assays and previous geological models. A simple labelling scheme was adopted assigning one label for an entire hole. The geology of the drilled area presented three classes of interest: shale, ore, and BIF. Table 2 presents the numerical results of this experiment.

Figure 2. Terex Reedrill SKSS-15 drill rig, used to collect data for the experiments.

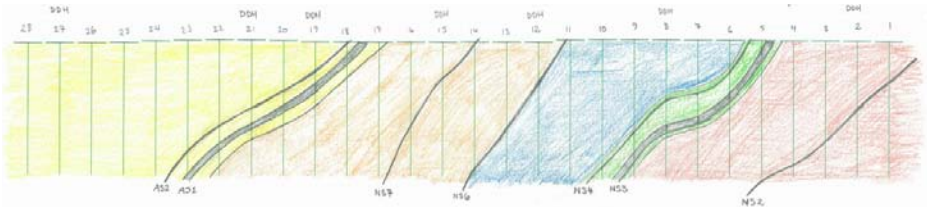
In our experiments, the CRFs using Logitboost outperformed the other approaches, providing better classification performance than LogitBoost alone. The algorithms performed well independent of the drilling mode. In the case of LogitBoost, the inputs are multiclass and the CRF has to simply associate neighbouring class labels. In the case of AdaBoost, the CRF also has to combine the several binary classifications provided by AdaBoost.

5. Conclusions

This work addresses the problem of incorporating spatial information into rock classification from MWD data. We have demonstrated the application of a CRF framework to estimate geology using drill sensor measurements. Experiments were conducted using real data collected by a production drill rig from an open pit iron ore mine. The CRF approach can provide an improvement in classification accuracy and F-measure—of more than 10% in some cases—compared to a simple boosting classifier. Drill performance measurements were classified into multiple classes according to the drilled rock characteristics.

Modelling spatial relationships is useful to exploit the fact that local lithology can be highly homogeneous. The proposed method associates data of nearby sections within individual boreholes. Future work will consider relationships between neighbouring boreholes.

(a) Original geophysical logging analysis



(b) Reference label set used to learn the models

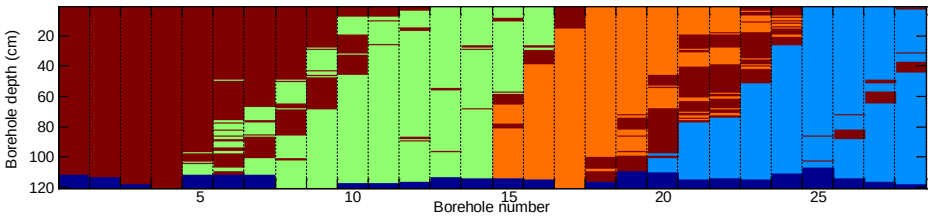


Figure 3. Geological interpretation of target region provided by mine geologists. Legend for (b) is shown in Fig. 4(c). BIF and shale units are waste rocks. Zone A and B are ore zones.

Table 1. Results for 28 boreholes classified into 4 categories

						Overall[d]	
		BIF	**ZoneA**	**ZoneB**	**Shale**	Micro	Macro
LB	A	94.51	87.70	89.14	81.95	76.65	88.32
	F	86.97	76.50	67.78	74.86	76.65	76.53
CRF[a]	A	94.66	**89.05**	90.52	82.58	78.41	89.20
	F	87.63	**79.29**	71.72	75.28	78.41	78.48
CRF[b]	A	**96.36**	83.99	90.40	78.84	74.80	87.40
	F	**91.28**	66.67	69.64	73.08	74.80	75.17
CRF[c]	A	96.20	86.41	**91.21**	**83.49**	**78.66**	**89.33**
	F	90.94	74.90	**73.48**	**76.68**	**78.66**	**79.00**

[a] CRF with LogitBoost
[b] CRF with AdaBoost (reset decision stumps)
[c] CRF with AdaBoost
[d] Overall performance was calculated by micro-averaging and macro-averaging for accuracy(A) and F-measure(F)

Table 2. Results for 132 boreholes classified into 3 categories

					Overall[d]	
		Shale	**Ore**	**BIF**	Micro	Macro
LB	A	88.84	82.99	88.55	80.19	86.79
	F	89.30	70.37	70.11	80.19	76.59
CRF[a]	A	**94.47**	**96.83**	**95.38**	**93.34**	**95.56**
	F	**94.70**	**94.40**	**88.20**	**93.34**	**92.43**
CRF[b]	A	86.27	81.82	88.85	78.47	85.65
	F	87.19	68.31	68.59	78.47	74.70
CRF[c]	A	91.23	88.85	93.73	86.90	91.27
	F	92.04	79.28	82.59	86.90	84.64

[a] CRF with LogitBoost
[b] CRF with AdaBoost (reset decision stumps)
[c] CRF with AdaBoost
[d] Overall performance was calculated by micro-averaging and macro-averaging for accuracy(A) and F-measure(F)

References

[1] W. H. Somerton, "A laboratory study of rock breakage by rotary drilling," *J. Petr. Tech., Trans. AIME*, vol. 216, pp. 92–97, 1959.

[2] R. Teale, "The concept of specific energy in rock drilling," *Int. J. Rock Mech. Min. Sci.*, vol. 2, pp. 57–73, 1965.

[3] R. L. King, M. A. Hicks, and S. P. Signer, "Using unsupervised learning for feature detection in a coal mine roof," *Eng. Appl. Art. Intell.*, vol. 6, no. 6, pp. 565–573, 1993.

[4] K. Itakura, K. Sato, Y. Ichihara, G. Deguchi, H. Matsumoto, and H. Eguchi, "Development of a roof logging system by rock bolt drilling," *Trans. Inst. Min. Metall. A*, vol. 106, pp. 118–123, 1997.

[5] W. K. Utt, "Neural network technology for strata strength characterization," in *Proc. Intl. Jt. Conf. Neur. Net.*, vol. 6, 1999, pp. 3806–3809.

[6] C. Sutton and A. Mccallum, "An introduction to conditional random fields for relational learning," in *Introduction to Statistical Relational Learning*, L. Getoor and B. Taskar, Eds. MIT Press, 2007, ch. 4.

[7] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. 13th Intl. Conf. Mach. Learn.*, 1996, pp. 148–156.

[8] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proc. 23rd Intl. Conf. Mach. Learn.*, 2006, pp. 161–168.

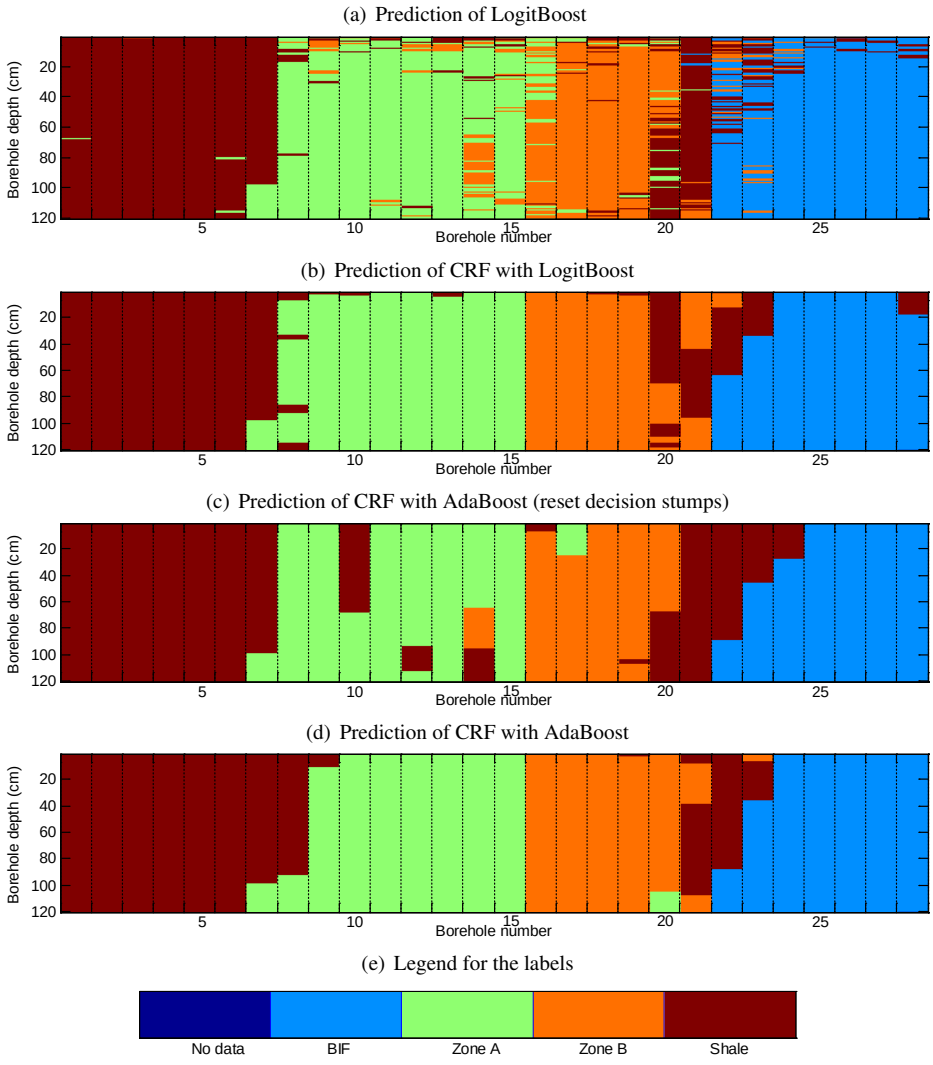


Figure 4. Predictions of the 28 boreholes' geological sections. Each borehole corresponds to a vertical row in the graph and the 3 m distance between holes is omitted.

[9] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Intl. Conf. Mach. Learn.*, 2001, pp. 282–289.

[10] J. Besag, "Statistical analysis of non-lattice data," *The Statist.*, vol. 24, no. 3, pp. 179–195, 1975.

[11] D. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Math. Programm.*, vol. 45, no. 3, Ser. B, pp. 503–528, 1989.

[12] S. Kumar and M. Hebert, "Discriminative random fields: A discriminative framework for contextual interaction in classification," in *Proc. IEEE Intl. Conf. Comp. Vis.*, vol. 2, 2003, pp. 1150–1157.

[13] M. Richardson and P. Domingos, "Markov logic networks," *Mach. Learn.*, vol. 62, no. 1–2, pp. 107–136, 2006.

[14] S. Friedman, D. Fox, and H. Pasula, "Voronoi random fields: Extracting the topological structure of indoor environments via place labeling," in *Proc. Int. Jt. Conf. Art. Intel.*, 2007, pp. 2109–2114.

[15] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Pub., 1988.

[16] P. Viola and M. Jones, "Robust real-time object detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[17] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *Ann. Stat.*, vol. 28, no. 2, pp. 337–407, 2000.