

Can Priors Be Trusted? Learning to Anticipate Roadworks

Bonolo Mathibela* and Michael A. Osborne† and Ingmar Posner* and Paul Newman*

Abstract—This paper addresses the question of how much a previously obtained map of a road environment should be trusted for vehicle localisation, during autonomous driving, by assessing the probability that roadworks are being traversed. We compare two formulations of a *roadwork prior*: one based on Gaussian Process (GP) Classification and the other a more conventional Hidden Markov Model (HMM) in order to model correlations between nearby parts of a vehicle trajectory. Importantly, our formulation allows this prior to be updated efficiently and repeatedly to gain an ever more accurate model of the environment over time. In the absence of, or in addition to, any *in-situ* observations, information from dedicated web resources can readily be incorporated into the framework. We evaluate our model using real data from an autonomous car and show that although the GP and HMM are roughly commensurate in terms of mapping roadworks, the GP provides a more powerful representation and lower prediction error. Our method allows us to truly map and anticipate roadworks on urban roads.

I. INTRODUCTION

Autonomous vehicles are expected to play a key role in future intelligent transport systems. Recent successful approaches to autonomous driving leverage prior maps of the environment traversed in order to engineer away some of the complexities of on-road navigation, environment interpretation and vehicle interaction (see, for example, [1], [2]). Such approaches inherently make the intuitive assumption that the layout of roads and related infrastructure is largely static. However, this can lead to potentially catastrophic failure when the information contained in the map is no longer correct. This is commonly the case when encountering roadworks: the road layout may change and temporary control infrastructure, such as traffic lights or stop signs, may be put in place. In 2009/2010, Transport for London (TfL) recorded about 370,000 roadworks carried out by London’s public utilities and highway authorities [3]. This figure is typical for most major cities and motivates the need to account for changing road structure. If autonomous vehicles are to function reliably in areas of road maintenance – either by taking appropriate action autonomously or by handing over control to a human operator – the correctness of the map prior must be assessed continuously, both *in-situ* and in trajectory planning. We envision a system catering for the continuous evolution of prior information using a multitude of information sources to be an essential stepping stone towards truly adaptive driver assistance systems and, ultimately, vehicle autonomy. This motivates our work.

The primary contribution of this paper is a framework for mapping roadworks, that seamlessly combines web-based resources with *in-situ* observations to update a *roadworks*



Fig. 1. This paper is concerned with the data fusion of web-based resources (websites citing the expected locations of roadworks) and in-situ observations (detections of roadwork indicators in Fig 2) to estimate a roadwork prior.

prior, queryable at any point along a trajectory. The assessment expressed by this prior is based on several sources of information: firstly, in the absence of any other information, an intuitive scepticism reflecting the general likelihood of roadworks; secondly, *in-situ* observations of features, such as signage or traffic cones, indicating the presence (or otherwise) of roadworks; and thirdly, information about ongoing roadworks provided by government agencies and disseminated via the Internet. We firstly employ a novel formulation of Gaussian Process (GP) classification, allowing for the use of noisy observations in order to account for correlations between observations from nearby points along the continuous trajectory. We compare this to a HMM formulation with continuous node observations. Both formulations enable life-long or multi-agent learning in that observations from multiple traversals of the same route are fused to continuously refine the prior until it converges to the ground truth. An overview of our proposed system is shown in Figure 1.

For the remainder of this paper we consider a scenario where a single vehicle traverses the same route multiple times. Following a review of related works in the next section, we describe how roadwork observations are acquired in Section III. Both generative models underlying our roadwork prior are introduced in Section IV and evaluated using real data in Section V. Conclusions are drawn in Section VI.

II. RELATED WORKS

The development of driver assistance systems including the detection and navigation of roadworks is an active area of research in the automotive industry (see, for example, the European Commission’s HAVEit project [4] or Volkswagen’s Temporary Autopilot (TAP) system). Much previous work has addressed the problem of detecting and classifying objects pertinent to autonomous driving such as obstacles on and alongside roads (see, for example, [5]). There also exists literature on the detection of road signs (see, for example, [6], [7]). However, the exploitation of prior maps of large-scale road networks for autonomous driving is a relatively recent development. To the best of our knowledge, the explicit modelling of the occurrence of roadworks based on a variety of information sources has not been addressed before.

* Mobile Robotics Group, University of Oxford

† Pattern Analysis Research Group, University of Oxford

{bonolo, ingmar, mosb, pnewman}@robots.ox.ac.uk



Fig. 2. The roadwork indicator objects that form the basis of our *in-situ* observations. We train classifiers to detect these objects.

III. ROADWORKS OBSERVATIONS

Various sources of information exist on which to base a model for the presence of roadworks at a given location. Here we briefly describe the two sources exploited in this work: *in-situ* observations obtained by an autonomous vehicle and derived from image data; and information from the Internet. The models derived in Section IV will be agnostic to the source of the information provided and will take as input simply *observations*. Let $y(x)$ represent the true binary label at a position x along a trajectory such that $y(x) = 1$ implies roadworks, $y(x) = -1$ implies not-roadworks. $\mathbf{z}(x)$ is the vector of roadwork observations, and Z is the matrix.

A. In-Situ Observations

In-situ observations are gathered by a vehicle as it travels along its trajectory. For the benefit of human (and non-human) drivers, roadwork sites are commonly marked by visible objects (roadwork indicators) examples of which are shown in Fig. 2. Let an *in-situ* observation, \mathbf{z}_{is} , be defined as a binary vector in an N -dimensional feature space such that $\mathbf{z}_{is} = [z_{is}^1, z_{is}^2, \dots, z_{is}^N]^T$. Each element z_{is}^i denotes the presence or absence of a feature derived from a particular roadwork indicator in a scene. In this work we exploit the presence or absence of any one of the roadworks indicators in Fig. 2.

Each image taken along a trajectory gives rise to an observation. Individual elements of the observation vector are set by applying dedicated object detectors to the associated image. The vision community has expended much effort to develop object detectors for a variety of objects pertinent to autonomous driving such as road signs. Here, we leverage a template matching approach as introduced in [8] within a detection framework based on GentleBoost [9]. In particular, a separate detector is trained for each one of the seven classes of common roadwork indicators listed in Fig. 2 using data gathered from the web. Observations are transformed into likelihood scores, $p(\mathbf{z}_{is}|y)$, using a naïve Bayes model which assumes that roadworks indicators occur independently from one another. Correlations between individual roadwork indicators are not directly modelled since the main focus of this work is the formulation of a *roadwork prior*. Thus,

$$p(\mathbf{z}_{is}|y) = \prod_{i=1}^N p(z_{is}^i|y). \quad (1)$$

B. Observations from Web Resources

Several online sources detailing locations of road disruptions exist. In this work, we harvest data from the Oxfordshire

county council website¹ as another source of observations. The website provides a comprehensive list of possible road disruptions, including roadworks. The *point* location, severity, estimated duration and type of road disruption is given. However, no indication of the extent or exact layout of the roadworks are provided – knowledge of which is essential if the use of a prior map is to be maximised.

In order to transform these observations into likelihood scores, $p(\mathbf{z}_{web}|y)$, for use in our model we need to assign a degree of trust in the correctness of the information provided. For this work we empirically set this trust to be complete:

$$p(\mathbf{z}_{web}|y = \text{roadworks}) = 1.0, \quad (2)$$

$$p(\mathbf{z}_{web}|y = \text{not-roadworks}) = 0.0. \quad (3)$$

It is worth noting that the trust placed in the city council information affects our predictions. This parameter can be tuned depending on the information source used; our faith in the city council may vary between municipalities and certainly from country to country. Alternatively, if desired, it could also be learned. Note also that even in this case of complete trust, given sufficient contrary evidence from our *in-situ* observations, our algorithm can recover from erroneous city council reports.

The roadworks model introduced in the next section does not differentiate between the source of an observation. For simplicity, in the remainder of the paper we will therefore drop the subscripts denoting whether an observation has been made *in-situ* or stems from a web resource.

IV. THE ROADWORKS PRIOR

We view roadwork detection as a problem of Bayesian inference. Our goal is to deduce the probability of roadworks being present at any point in continuous space and time, given *in-situ* noisy observations and uncertain data from the web. To model this, we consider both a Gaussian Process and HMM roadworks function to allow us to compute the posterior probability of any point belonging to roadworks.

A. Hidden Markov Model for Roadwork Mapping

Hidden Markov Models (HMMs) provide a powerful framework for probabilistically modelling sequential data. A HMM is comprised of a set of variables for T discrete steps, $\mathbf{y}_{1:T}$ (representing the vector of latent variables) and $Z'_{1:T}$ (representing the matrix of observations) that adhere to predefined conditional independence assumptions. These assumptions can be summarised by the joint distribution over all variables given by:

$$p(Z'_{1:T}, \mathbf{y}_{1:T}) = p(y_1) \prod_{t=2}^T p(y_t | y_{t-1}) \prod_{t=1}^T p(\mathbf{z}'_t | y_t) \quad (4)$$

The prior distribution $p(y_1 = i) = \pi_i$ for each state i (roadworks or not-roadworks) expresses our initial belief in the state distribution of the first node of the HMM. π_i form hyperparameters of the HMM, along with the state transitional probabilities $p(y_t = i | y_{t-1} = j) = G_{i,j}$ (note: our HMM is

¹<http://voyager.oxfordshire.gov.uk/network.aspx>

time-homogenous therefore transitional probabilities are time-independent). Finally, we model the observation probability distributions as Gaussians with mean α_j and covariance A_j ,

$$p(\mathbf{z}'_t | y_t = j) = \mathcal{N}(\mathbf{z}'_t; \alpha_j, A_j). \quad (5)$$

Hyperparameters $\pi_i, G_{i,j}, \alpha_j$ and A_j are set to the maximum *a posteriori* estimates of the parameters using Expectation Maximization [10]. We train the HMM in batch mode, using data Z from both *in-situ* and city council observations. Our observation matrix $Z'_{1:T}$ comprises the likelihoods of both roadworks and not roadworks, each row of which is a binary vector $\mathbf{z}'_t = (p(\mathbf{z}_t | y_t = -1), p(\mathbf{z}_t | y_t = 1))$.

B. Gaussian Process Classification for Roadwork Mapping

Gaussian processes (GPs) [11] provide a powerful framework for performing Bayesian inference about functions. A Gaussian process is a distribution on the functions $f: \mathcal{X} \rightarrow \mathbb{R}$ (on any arbitrary domain \mathcal{X}) such that the distribution of the function values at a finite subset of points $F \subseteq \mathcal{X}$ is a multivariate Gaussian. A Gaussian process is completely defined by its mean function $\mu: \mathcal{X} \rightarrow \mathbb{R}$ and a positive semidefinite covariance function $K: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$. The mean function describes what we would expect for f before receiving any data and the covariance function describes how function values are correlated as a function of their locations.

Given our initial ignorance about the locations of roadworks, we set the mean function to an (unknown) constant μ . To express our expectation that our roadworks function be smooth, we choose the squared exponential covariance,

$$K(x_1, x_2; \lambda, \Sigma) = \lambda^2 \exp\left(-\frac{1}{2}(x_1 - x_2)^\top \Sigma^{-1} (x_1 - x_2)\right). \quad (6)$$

Here λ specifies an output scale for f , controlling the magnitude of its variation. Since there is little inter-dependence amongst our inputs, we choose Σ as a diagonal matrix whose diagonal elements represent squared input scales, controlling the relative scales of variation in each dimension of x . The parameters of our mean and covariance functions, μ, λ and Σ form elements of the vector θ , the hyperparameters of our model. Given data, these hyperparameters of the GP will be selected as those that return the maximal marginal likelihood.

Given a Gaussian process prior distribution for the function $f: \mathcal{X} \rightarrow \mathbb{R}$, and a set of input points \mathbf{x} , the prior distribution on $\mathbf{f} = f(\mathbf{x})$ is

$$p(\mathbf{f} | \mathbf{x}, \theta) = \mathcal{N}(\mathbf{f}; \mu(\mathbf{x}; \theta), K(\mathbf{x}, \mathbf{x}; \theta)), \quad (7)$$

where $K(\mathbf{x}, \mathbf{x}; \theta)$ is the Gram matrix of the points \mathbf{x} and \mathbf{f} is a real-valued latent variable. For traditional, non-noisy, classification we observe $y \in \{-1, 1\}$ binary-valued class labels. Our observation likelihood, combined with i.i.d. Gaussian noise in f , is given by the cumulative Gaussian observation likelihood

$$p(y | f, x, \sigma^2) = \Phi(yf; 0, \sigma^2), \quad (8)$$

$$\Phi(a; \mu, C) = \int_{-\infty}^a \mathcal{N}(a'; \mu, C) da'. \quad (9)$$

For our application, however, our observations are actually of a quantity \mathbf{z} that does not definitively specify a class



Fig. 3. The Wildcat autonomous car, pictured above, was used to collect the data used in this paper. It is equipped with a Bumblebee stereo camera, capturing 512x384 colour images at 20Hz and a differential GPS/IMU which we use for localizing city council observations along the vehicle's trajectory. Data was collected over a period of 3 months. In this paper, we process 38.98GB of data.

label. Distinct from some previous work on GP classification with noisy labels [12], [13], some observations are more indicative of the label than others; the uncertainty varies with observation. Other approaches [14], [15], [16] consider the case in which a binary class label is observed, which may have been flipped with some probability; we instead observe only the presence or absence of features as indirect evidence of our class label. For our application, we propose an observation likelihood

$$\begin{aligned} p(\mathbf{z} | f, x, \sigma^2) &= \sum_{y \in \{-1, 1\}} p(\mathbf{z} | y, x, \sigma^2) p(y | f, x, \sigma^2) \\ &= \sum_{y \in \{-1, 1\}} p(\mathbf{z} | y, x, \sigma^2) \Phi(yf; 0, \sigma^2). \end{aligned} \quad (10)$$

Our non-Gaussian observation likelihood (10) led us to use Expectation Propagation [17] in order to perform inference. Expectation propagation approximates our likelihoods (10) as Gaussians in f , and uses sequential moment-matching to fit the parameters of each one. Results for both of our models are discussed in the next Section. While we performed our experiments in batch-mode, efficient GP updates exist to enable on-line experiments.

V. EXPERIMENTAL RESULTS

Our roadworks prior was evaluated using image and GPS data gathered over several kilometres of track with our autonomous vehicle, a Bowler Wildcat (see Figure 3). The data collected spans approximately 5km of track covering roadworks (in both urban and rural areas) classified using detectors described in Section III-A. HMM classification [18], [19] was implemented using the PMTK toolbox [20]. Similarly, GP classification was implemented using the GPML toolbox [21] with a likelihood function reflecting (10).

We evaluate our methods according to two metrics, both averages over M locations, indexed by i , and N repeated trials, indexed by j . Define $\tilde{y}_{i,j}$ as the ground truth state for the i th location in the j th experiment, and $p(y_{i,j} | Z_j)$ is the posterior produced by a method for those states given observation matrix Z_j . Firstly, we computed the root mean normalised square error,

$$\text{RMSE} = \left(\frac{1}{M N} \sum_{i=1}^M \sum_{j=1}^N (\tilde{y}_{i,j} - p(y_{i,j} | Z_j))^2 \right)^{-\frac{1}{2}}, \quad (11)$$

a measure of the accuracy of our probabilistic estimates. A more sophisticated metric is the negative log-density of the truth, assuming experiments are independent,

$$\text{NLL} = - \sum_j^N \log p(y_{i,j} = \tilde{y}_{i,j}, \forall i = 1, \dots, M | Z_j). \quad (12)$$

A. Synthesised Excursions

Using *real* images gathered with our autonomous vehicle, we set up synthesised excursions by defining a fixed road interval containing a unique ground-truth roadworks configuration. For this road interval, we allow our algorithm two points at which roadworks have been identified, simulating the county council prior. To simulate the information gleaned by our vehicle, we randomly sample a sequence of 50 *real* images, consistent with the specified roadworks layout, to create a single traversal. This process is repeated for 10 traversals. We proceed in this manner to simulate 15 independent excursions.

During each traversal, the *in-situ* observations are added to both the Gaussian process and the HMM-based model. Figures 4 and 5 show the posterior of roadworks generated by each model respectively for the same sample excursion as the number of traversals increases. In an ideal case, the probability of roadworks at any point would match the ground-truth. These figures suggest that both algorithms become more confident about the presence or absence of roadworks as the number of traversals increases. Both successfully discount the erroneous web-prior. We quantify this convergence using the root-mean-square error (RMSE) between the roadworks posterior and the binary-valued function indicating the true presence of roadworks. Figure 6 shows how, on average, the RMSE tends to decrease with increasing number of traversals. Crucially, however, while the RMSE median at each traversal is roughly commensurate for both approaches, the GP-based model tends to exhibit a much reduced variance compared to the HMM. This indicates that the GP-based model performs well over the full range of scenarios presented, whereas the HMM-based predictions can be substantially inferior. We attribute this to the smoothness assumption inherent in our GP which allow the system to learn a more flexible model as well as the usual spatial extent of roadworks. Figure 7 shows that the likelihood of the GP model substantially improves with increasing traversal, whereas the HMM is hindered by its poor performance on outliers.

B. Repeated Traversals

In addition to the synthesised excursions we repeatedly traversed a real roadworks site around half a kilometre in extent. The GP-based model output for these data is shown in Figure 8. Note that in this real scenario observations may not coincide spatially with previous readings and thus would require a more involved continuous HMM formulation. Our GP-based model, on the other hand, naturally accommodates continuous space. We do not here explore a continuous HMM formulation. As in the simulations, inspection of Figure 8 reveals that the peaks in the predictions become increasingly correlated with the detailed structure of the ground-truth. The effect here is less pronounced than in the simulations

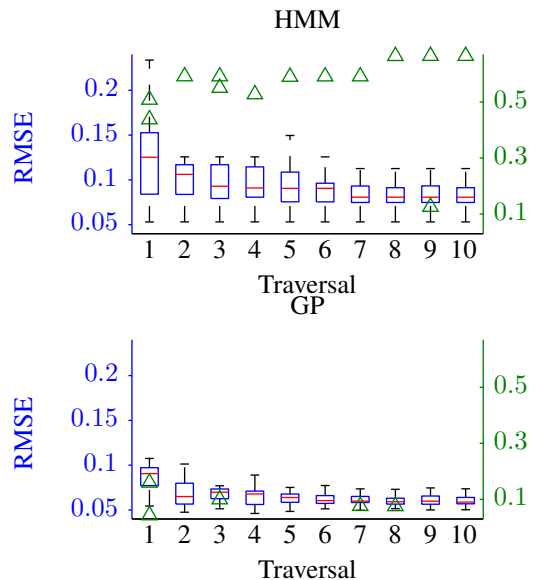


Fig. 6. Box plots of the RMSE error for the HMM and the GP *roadwork* priors across the same 15 simulated excursions, each containing 10 traversals. The plots are a non-parametric way of showing the sample minimum, 25th percentile, median, 75th percentile, maximum and samples considered outliers (plotted as triangles with right hand y-axis). Outliers are points that lie beyond the boxplot fence and signify extreme values in the tails of the distribution. On average, the RMSE is decreasing for both plots. The HMM plot has significantly more outliers than the GP and a higher variance, which implies that the GP performs better across the entire range of scenarios presented.

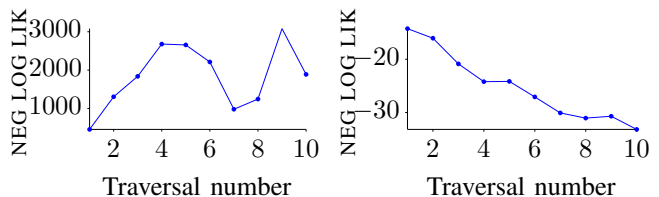


Fig. 7. Negative log likelihoods for HMM, left, and GP, right, over the simulated traversals. Note that while the different formats of data used by the two models imply that their likelihoods are not directly comparable, the plots do illustrate that the NLL of the GP exhibits a strong downward trend; the GP is increasingly accurate with increasing traversal number.

due to an increased amount of contradictory information (mis-classifications of roadworks indicators). This can be attributed to environmental factors (e.g. lighting) remaining constant during real traversals and thus potentially introducing a systematic bias to results. This bias is not captured by our simulator, which randomly samples real scenes from a variety of traversals. However, even in this more challenging scenario our GP-based model accurately maps the roadworks. Also note that the inaccurate web-prior is successfully discounted.

VI. CONCLUSIONS AND FUTURE WORK

This work proposes two possible probabilistic models for a *roadwork prior* – that is, a prior expressing for each position along a vehicle’s trajectory the probability of encountering roadworks. Our models are built using observations taken both *in-situ* and obtained from dedicated web resources. We consider both GP classification and an HMM-based formulation for this prior. In the GP case, the *roadwork prior* is based on a novel formulation of GP classification which accommodates

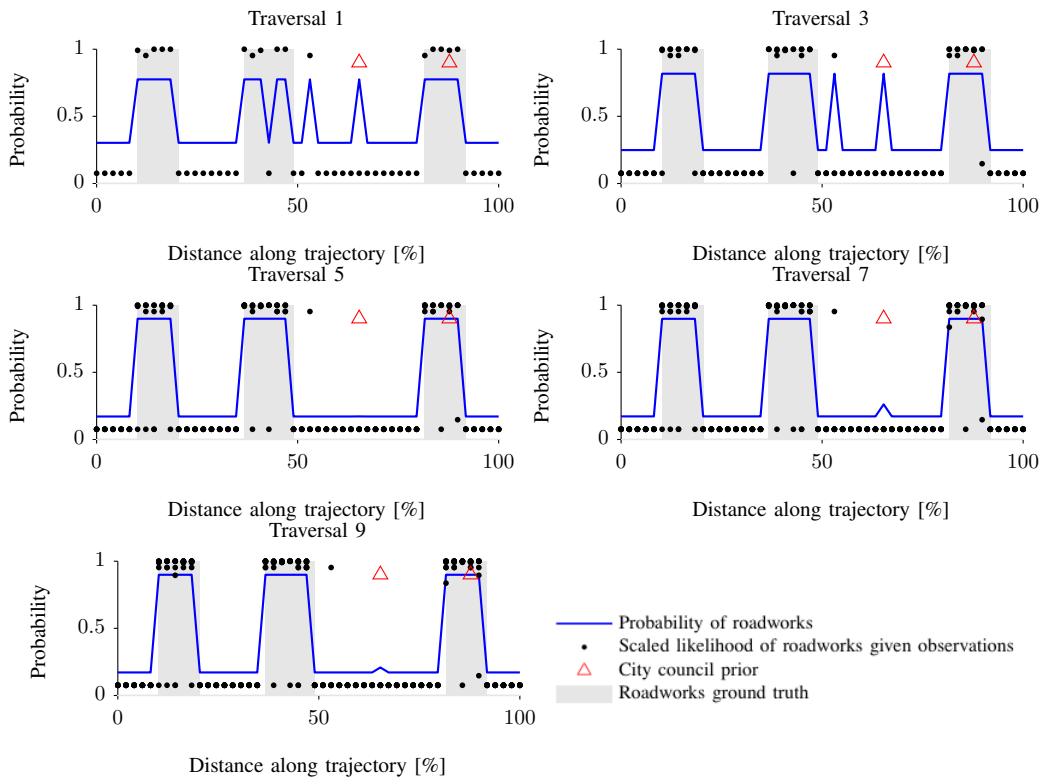


Fig. 4. The *roadwork prior* estimated by the HMM is shown for every second traversal. Notice how this prior converges to the groundtruth with increasing number of traversals. The erroneous city council prior is also successfully ignored from the 5th traversal onwards. We compare the exact same roadwork configuration as in Fig. 5 Note that the plotted likelihoods are scaled by a multiplicative factor so that the sum of the likelihood of roadworks and the likelihood of not-roadworks is equal to unity for every point; roughly speaking, the posterior should follow these scaled likelihoods.

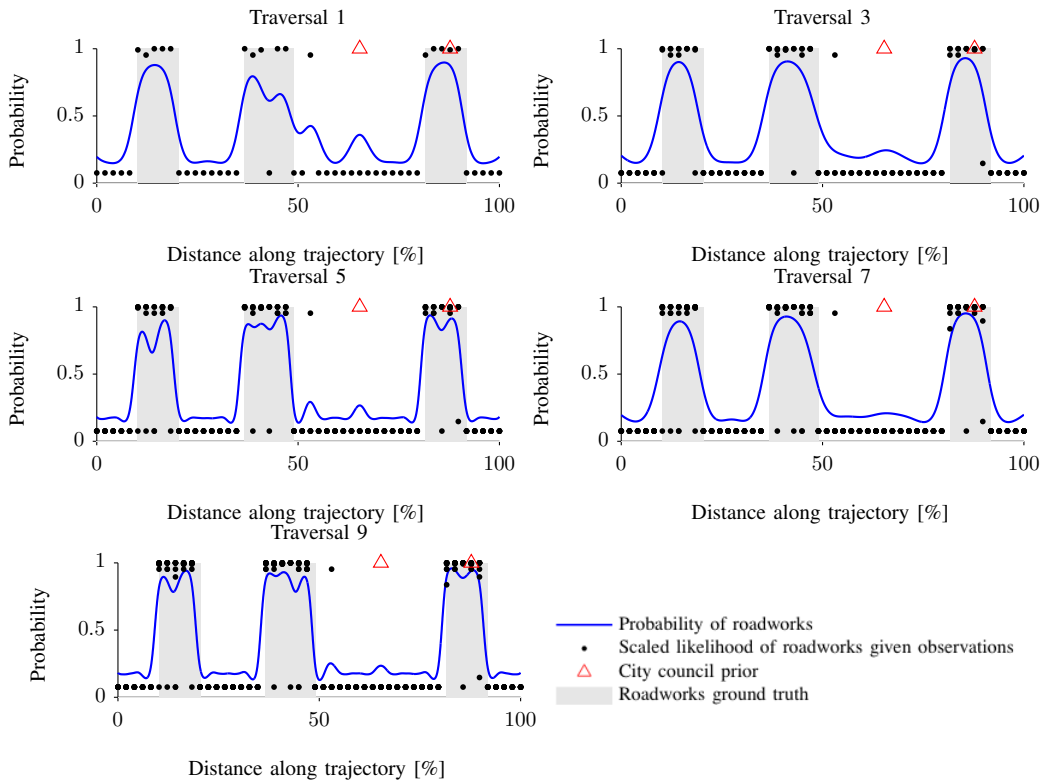


Fig. 5. The *roadwork prior* estimated by the GP is shown for every second traversal. Notice how this prior converges to the groundtruth with increasing number of traversals. The erroneous first city council prior is also successfully ignored from the 3rd traversal onwards. Compared to the HMM implementation (see Fig. 4), the GP *roadwork prior* is inherently smoother demonstrating realistic progressions from roadwork to non-roadwork states and vice versa. Note that the plotted likelihoods are scaled similar to Fig. 4.

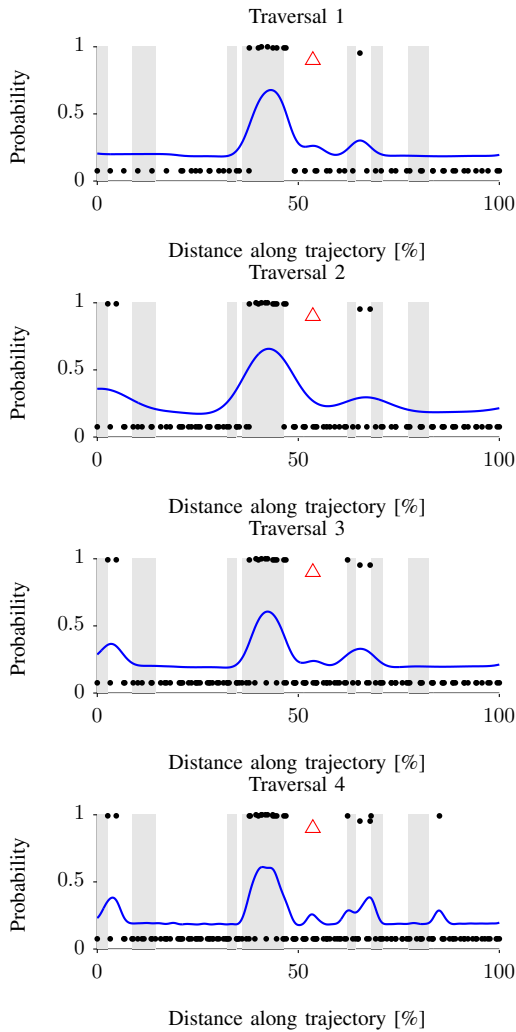


Fig. 8. For our real excursions (600m long), the Oxfordshire county council provided a single observation and our GP framework successfully recovered the true roadwork layout. The sparsity of the groundtruth is due to the nature of the roadwork scene: roadwork signs leading up to some barriers. Furthermore, for real traversals, observations occur in continuous space and are unlikely to exactly coincide (as in the case for simulations). In spite of this, our system is able to accurately predict the underlying roadwork prior for this real roadwork scene. For legend, see Fig. 5.

noisy observations in a continuous state space. For both models, repeated updates are possible, allowing an agent to refine its predictions over time. In particular, we demonstrate that, while both approaches converge towards the groundtruth with increasing traversals, only our proposed GP model performs well across the entire range of scenarios presented.

There are several avenues for future, related work. Thus far, we have considered functions of a univariate spatial x . More interestingly, we could add in a second spatial dimension to allow for inference over two dimensional road networks. A temporal dimension for explicitly modelling the temporary nature of roadworks would also permit dynamic inference.

VII. ACKNOWLEDGEMENTS

Michael A. Osborne was funded by the Systems Engineering for Autonomous Systems (SEAS) Defence Technology

Centre established by the UK Ministry of Defence. Paul Newman was supported by an EPSRC Leadership Fellowship, EPSRC Grant EP/I005021/1. Bonolo Mathibela was funded by the Rhodes Trust. This work has partly been supported by the EU Project EUROPA-FP7-231888. The authors are grateful to Dr Benjamin Davis for assisting with data collection and anonymous reviewers for helpful comments.

REFERENCES

- [1] N. Fairfield and C. Urmson, "Traffic light mapping and detection," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, may 2011, pp. 5421–5426.
- [2] J. Levinson, J. Askeland, J. Dolson, and S. Thrun, "Traffic light mapping, localization, and state detection for autonomous vehicles," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, may 2011, pp. 5784–5791.
- [3] S. T. Panel, "Minimising disruption from roadworks," Transport for London, Tech. Rep., may 2010.
- [4] European Commission, "HAVEit Project," <http://www.haveit-eu.org/>, 2011.
- [5] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, K. Lau, C. Oakley, M. Palatucci, V. Pratt, P. Stang, S. Strohband, C. Dupont, L.-E. Jendrossek, C. Koelen, C. Markey, C. Rummel, J. van Niekerk, E. Jensen, P. Alessandrini, G. Bradski, B. Davies, S. Ettinger, A. Kaehler, A. Nefian, and P. Mahoney, "Stanley: The robot that won the DARPA Grand Challenge," *Journal of Field Robotics*, vol. 23, no. 9, pp. 661–692, 2006.
- [6] R. Timofte, K. Zimmermann, and L. V. Gool, "Multi-view traffic sign detection, recognition, and 3D localisation," in *Applications of Computer Vision (WACV), 2009 Workshop on*, dec. 2009, pp. 1–8.
- [7] A. Broggi, P. Cerri, P. Medici, P. Porta, and G. Ghisio, "Real time road signs recognition," in *Intelligent Vehicles Symposium, 2007 IEEE*, june 2007, pp. 981–986.
- [8] A. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing features: Efficient boosting procedures for multiclass object detection," in *Computer Vision and Pattern Recognition*, 2004, pp. 762–769.
- [9] R. T. J. Friedman, T. Hastie, "Additive logistic regression: a statistical view of boosting," *The Annals of Statistics*, vol. 28, pp. 337–407, 2000.
- [10] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1–38, 1977.
- [11] C. Rasmussen and C. Williams, *Gaussian processes for machine learning*. MIT Press, 2005.
- [12] M. Opper and O. Winther, "Mean field methods for classification with gaussian processes," in *Advances in neural information processing systems 11: proceedings of the 1998 conference*, vol. 11. The MIT Press, 1999, p. 309.
- [13] S. Yu, B. Krishnapuram, R. Rosales, H. Steck, and R. Rao, "Bayesian co-training," *Advances in neural information processing systems*, vol. 20, pp. 1665–1672, 2008.
- [14] A. Kapoor, "Learning discriminative models with incomplete data," Ph.D. dissertation, Massachusetts Institute of Technology, 2006.
- [15] C. Pal, G. Mann, and R. Minerich, "Putting semantic information extraction on the map: noisy label models for fact extraction," in *Proceedings of the Workshop on Information Integration on the Web at AAAI*, 2007.
- [16] Y. Xue, D. Williams, and H. Qiu, "Classification with imperfect labels for fault prediction," in *Proceedings of the First International Workshop on Data Mining for Service and Maintenance*. ACM, 2011, pp. 12–16.
- [17] T. Minka, "Expectation propagation for approximate Bayesian inference," in *Uncertainty in Artificial Intelligence*, vol. 17. Citeseer, 2001, pp. 362–369.
- [18] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [19] J. Bilmes, "What HMMs can do," Department of Electrical Engineering, University of Washington, Tech. Rep. UWEETR-2002-0003, 2002.
- [20] K. Murphy and M. Dunham, "PMTK: Probabilistic modeling toolkit," in *Neural Information Processing Systems (NIPS) Workshop on Probabilistic Programming*, 2008.
- [21] C. Rasmussen and H. Nickisch, "Gaussian processes for machine learning (GPML) toolbox," *The Journal of Machine Learning Research*, vol. 11, pp. 3011–3015, 2010.