Reading the Road: Road Marking Classification and Interpretation

Bonolo Mathibela, Paul Newman, and Ingmar Posner

Abstract—Road markings embody the rules of the road whilst capturing the upcoming road layout. These rules are diligently studied and applied to driving situations by human drivers who have read Highway Traffic driving manuals (road marking interpretation). An autonomous vehicle must however be taught to read the road, as a human might. This paper addresses the problem of automatically reading the rules encoded in road markings, by classifying them into seven distinct classes: single boundary, double boundary, separator, zig-zag, intersection, boxed junction and special lane. Our method employs a unique set of geometric feature functions within a probabilistic RUSBoost and Conditional Random Field (CRF) classification framework. This allows us to jointly classify extracted road markings. Furthermore, we infer the semantics of road scenes (pedestrian approaches and no drive regions) based on marking classification results. Finally, our algorithms are evaluated on a large real-life ground truth annotated dataset from our vehicle.

Index Terms—Road marking classification, scene interpretation, scene understanding, situational awareness.

Fig. 1. Reading the road requires classifying and interpreting the meaning of road markings in the road scene. Here, each colour represents a unique road marking class.

I. INTRODUCTION

FOR autonomous driving to occur, there needs to be some degree of situational awareness on the part of the robotic car, allowing it to safely navigate its environment. Formally, situational awareness is defined as a three level process involving "the perception of elements in the environment within a volume of time and space, and the comprehension of their meaning, and the projection of their status in the near future" [1]. In the context of autonomous driving, the ultimate goal of situational awareness is to assist in informative decision making. We approach situational awareness in terms of interpreting the underlying meaning of road markings, thus allowing a vehicle to *read the road* ahead, as shown in Fig. 1. Here, reading the road is an online process of classifying and interpreting road markings along unknown roadways.

This greatly contrasts offline approaches (such as [2] and [3]) where a semantic prior map (of road markings for example) is first created by surveying known roadways and a vehicle localises against this map when traversing the same known roadway. The two main disadvantages of the offline approach

is that the vehicle cannot interpret a new area without first creating an offline semantic map of that area; and roads are changing due to maintenance, construction and roadworks [4], [5]. Although roadworks are temporary, their frequency of occurrence is surprisingly high. In 2009/2010 alone, Transport for London reported an estimated 370 000 roadworks [6], a figure that is typical for most major cities.

Highway authorities worldwide publish technical reports detailing meanings and placements of road markings to aid human drivers in reading the road scene ahead. A driver can predict upcoming road topology, know the set of permissible manoeuvres, and keep to the correct lane, based on interpreting road markings. An autonomous vehicle must however be taught to automatically read and interpret road semantics as a human might—this motivates our work.

This paper extends our previous work [7] and makes three contributions. Firstly, we propose a new probabilistic framework for classifying and interpreting road markings in images. Our framework accommodates a varying graphical model structure (each road scene gives rise to a unique structure). Secondly, we create a benchmarking ground truth class annotated dataset containing 2068 images and over 13 099 unique annotations, the largest of its kind to date, to the best of our knowledge. Large road marking datasets are hard to come by, apart from [8] which does not contain class labels (and has just 116 images) and [9] which has bounding box ground truth labels of painted road signs in 1208 images. Thirdly, we show how traffic-light controlled pedestrian crossings and no-drive regions can be automatically detected from classified road markings.

Following a review of related works in the next section, we discuss our representation in Section III, our road marking

classification framework in Section IV, evaluations in Section V, and finally conclude in Section VI.

II. RELATED WORKS

Road marking classification and interpretation is a problem that is of great interest to the intelligent transportation community. In earlier work, Charbonnier *et al.* extracted parameters such as dash modulation and line width by representing lane markings as polygons of grouped detections; and could also detect arrow type road markings [10].

The simplest approaches to road marking classification use template matching, such as [11] for the recognition of crosswalks and arrows using single or repetitive rectangular pattern templates, and [12] for detecting arrow markings. Template matching approaches generally perform poorly when the road markings are partially observed or occluded.

Descriptor and Fourier-analysis based approaches have also been proposed before. Wu and Ranganathan's HOG-based (Histograms of Oriented Gradients) [13] road marking speed sign, arrow and word detection [9] algorithm uses Feature matching, followed by structural matching to capture the 2D geometry of FAST interest points. Collado *et al.'s* lane interpretation system for classifying continuous, discontinuous and merge type lanes [14] uses Fourier analysis on the intensity line profile of thresholded images for classification. Thresholds are applied to the resulting Fourier power spectrums within heuristically determined frequency bands. In later work [15] identify one-way, two-way or freeway roads based on Fourier analysis of road markings to classify them as continuous, broken or merge lines.

Prior geometric information has also been exploited before: zebra crossings and dashed lines were reconstructed in [16], where known 3D size and shape information of dashed lines and zebra crossing road markings was exploited. A signature-based recognition approach exploiting the alignment of edges with the road axis was used: projections of the edges along horizontal and longitudinal axes provided accumulation signature features [16] (e.g., repetitive peaks for zebra crossings versus a pair of peaks for dashed lines in the transverse axis), where the road axis is taken to be the most frequently occurring orientation of 2D line edges. The system was tested on just 1km of data (150 stereo pair images).

In recent years, classifier based approaches have become popular, and examples include a decision tree classifier for intersection road markings trained on six numerical features based on relationships between the road marking border points and an extracted RANSAC line fit [17], moments to capture shapes of arrows and straight lines [18], and an Artificial Neural Network (ANN) classifier with invariant spatial moments, histogram projections and normalised angular measurement features for recognising six road marking arrow types and 17 alpha numeric characters [19]. Finally, a generative learning method based on eigenvectors was used by [20] to recognise arrow and numbered markings.

In the most related work, Duchow and Kortner detect individual lane segments and aggregate these into intersection lanes [21]. A machine learning approach is proposed where a set of

Fig. 2. Zig-zag approach road markings lead up to a traffic light controlled pedestrian crossing.

features (width, length, orientation, position, intensity differences and deviations) is extracted from single, pair and triplet groups of lane marking segments and a Support Vector Machine (SVM) is trained to classify lane segments as belonging to the same lane or not [21]. Triplets are then combined to form lane boundaries. Importantly, this approach requires no use of prior information (digital maps or geometric road models) and neighbourhood information is captured by considering triplets of road markings at a time. Although [21] considers only intersection classes, their work is the most related to ours as relationships between neighbouring road marking segments are exploited by the classification framework and no explicit geometric priors are used, similar to our approach.

Unlike previous work which focuses on finding arrows and painted traffic signs on roads [9], [12], [17]–[20], classifying solid and dashed lines [14], intersections [21] and zebra crossings [11], [16], to the best of our knowledge, we are the first to address the problem of road marking joint classification (where markings are classified together, relative to each other) and interpretation (deriving the semantics of road scenes from road markings) for situational awareness. Furthermore, with available data, additional classes can readily be incorporated into our classification framework.

III. REPRESENTATION

Our goal is for a vehicle to read and interpret the rules of the road encoded in road markings in order to gain a sense of situational awareness of the current and upcoming road topology [7]. For example, observing zig-zag approach road markings informs a vehicle that there is an upcoming traffic light controlled pedestrian crossing (see Fig. 2). Successful recognition of zig-zag approaches therefore provides strong priors for the location of traffic lights and pedestrians, which is useful also for cuing the appropriate detector modules.

A. Seven Road Marking Classes

We classify the seven types of road markings shown in Table I and the accompanying figure. Here, we observe that:

- The number of lanes influences the type of road markings present. We presently do not rely on having prior knowledge of the number of lanes present; but in future work this is a parameter we will learn from combining local vehicle observations with web-based sources.
- The class of a specific road marking is strongly influenced by that of its neighbours (notice how similar

TABLE I
WE CLASSIFY SEVEN DISTINCT ROAD MARKING TYPES (CLASSES)
MOST COMMONLY FOUND ON BRITISH ROADS. ROAD MARKING
SEMANTIC CATEGORIES VARY BETWEEN COUNTRIES: THE USA, U.K.
AND SOUTH AFRICA HAVE OVER 24, 29, AND 22 TYPES RESPECTIVELY

Class	Description	Abbreviation	Colour
1	Single boundary	SGL	Red
2	Double Boundary	DBL	Green
3	Separator	DIV	Magenta
4	Zig-zag	ZIG	Cyan
5	Intersection	INT	Yellow
6	Boxed Junction	BJ	Blue
7	Special Lane	SPL	Orange

colours/classes are clustered). Thus we model the relationships between road markings based on proximity.

- The perspective effect in the images distorts the true lengths, orientations and positions of road markings. To attempt to recover the true geometric properties of road markings, we make a "flat-earth" assumption and assume a fixed known height for the camera relative to the ground plane. Although the assumption of a locally planar road performs satisfactorily, it does not strictly hold when changes in elevation and turning occur. Algorithms such as Fast Inverse Perspective Mapping Algorithm (FIPMA) adaptively update the camera parameters [22] to account for these subtle changes.

B. Inverse Perspective Mapping (IPM)

As a pre-processing step, we begin by removing the perspective effect from images using Inverse Perspective Mapping [23]. Perspective in road images distorts true lengths, positions and orientations of road markings as shown in Fig. 3. This effect is especially pronounced in zig-zag type markings (cyan colour) where the lengths of each road marking segment appear to be different when, in reality, they are the same length and one of two opposing orientations. To address this, the linear perspective can either be implicitly modeled [10], [17] or Inverse Perspective Mapping (IPM) can be performed [9], [18]–[21], [24] to create a perspective free bird's-eye-view image. We adopt the latter approach and compute a homography matrix $\mathbf{H} = \mathbf{KRK}^{-1}$, which defines the one-to-one relations between the camera image pixels and synthesised bird's-eye-view pixels. Here, $\mathbf{R}$ is the virtual rotation matrix and $\mathbf{K}$ is the camera's intrinsic parameters.

Original Images

Inverse Perspective Transformed Images

Fig. 3. Inverse Perspective Mapping (IPM) to remove the perspective effect in images. Notice how, in the IPM images, parallel lines no longer intersect at the vanishing point. Thus true lengths, orientations and positions can be recovered.
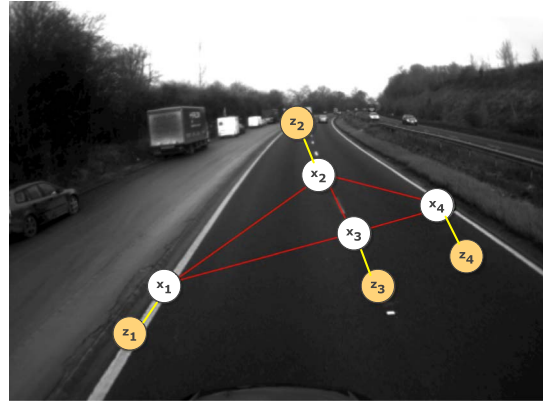
Fig. 4. Our CRF model for road marking classification superimposed on the road image. The hidden states $\mathbf{x_i}$ are the unknown road marking classes and the observations $\mathbf{z_i}$ (orange) correspond to line segment endpoints extracted from each road marking. This graph is a probability distribution $p(\mathbf{x}|\mathbf{z})$ and each road image example has a unique graph.

C. Line Segment Representation

Assuming road markings have already been detected, a geometric model is assigned to each marking based on the system application. For lane detection and tracking applications, smoothness and continuity is an important consideration, thus spline [25] based geometric models are often used. We choose a line segment geometric model similar to [14], [21]. A painted road marking segment is therefore represented by a line segment (or series of line segments) parameterised by two endpoints p_A and p_B, where the midpoint locations, $p_C \triangleq (p_A + p_B)/2$, represent node locations for our road marking connectivity graphical structure. For solid continuous road markings extending beyond the image capture area, we assume that the marking endpoints coincide with the image edge. Curved road markings are represented using a series of line segments thus creating a piece-wise approximation (see Fig. 1). We therefore frame road marking classification in terms of assigning the most likely road marking class to each node.

IV. ROAD MARKING CLASSIFICATION

A natural way to represent road markings is to use a graph, where the nodes represent observations (line segments) and hidden states (unknown road marking class assignment), and the edges represent the relational connectivity (see Fig. 4).
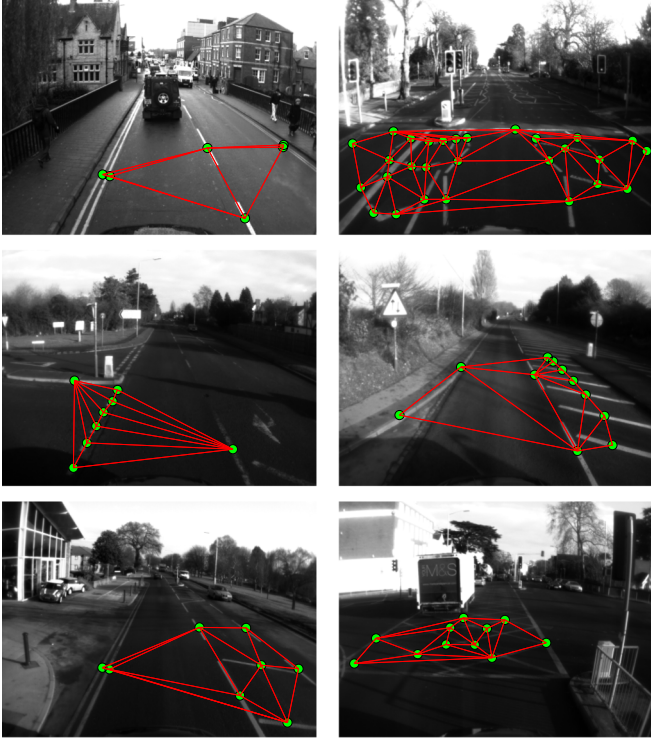
Fig. 5. Road marking graph structures created using Delaunay Triangulation. Notice how each example is unique.

A. Graphical Structure

The graph structure (connectivity) captures relationships between road marking segments. In the Figure of Table I, we observe that road markings belonging to the same class tend to be spatially clustered into groups (notice how the same colours in the image appear clustered). This implies that adjacent road markings are likely to belong to the same class, thus nodes should be connected to their closest neighbours. We therefore use Delaunay triangulation (the nearest neighbour graph is a subgraph of Delaunay Triangulation [26]) to create a graph structure of road markings, assuming a locally planar road surface. Here, a unique graph structure exists per training example as illustrated by Fig. 5. Given the graphical structure we can now model the road using a Conditional Random Field.

B. Conditional Random Fields

Conditional Random Fields (CRFs) are undirected probabilistic graphical models for relational learning, initially proposed by Lafferty *et al.* [27]. CRFs directly model the conditional distribution over hidden variables $\mathbf{x}$, given observations $\mathbf{z}$ as $p(\mathbf{x}|\mathbf{z})$. In the graphical model, each node represents the hidden states and observations. Thus, given a set of n hidden states $\mathbf{x} = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}\}$ and data $\mathbf{z}$, the nodes $\mathbf{x_i}$ and their associated graph structure define the conditional distribution $p(\mathbf{x}|\mathbf{z})$ over the hidden states (see Fig. 4). Each fully connected subgraph of this graph structure represents a clique in the graph. Let C be the set of cliques in the graph, then a CRF factorizes the conditional distribution $p(\mathbf{x}|\mathbf{z})$ into a product of clique potentials $\phi_c(\mathbf{z}, \mathbf{x_c})$ where $c \in C$, and $\mathbf{x_c}$ and $\mathbf{z}$ are the hidden nodes and data of clique c, respectively. Each

clique potential $\phi_c(\mathbf{z}, \mathbf{x_c})$ is a strictly positive function of the variable configuration and captures the compatibility amongst variables in the clique (thus large values imply a more likely configuration) [28]. The Conditional Random Field model can therefore be written as:

$$p(\mathbf{x}|\mathbf{z}) = \frac{1}{Z(\mathbf{z})} \prod_{c \in C} \phi_c(\mathbf{z}, \mathbf{x_c}) \tag{1}$$

where the normalisation factor (partition function) is $Z(\mathbf{z}) = \sum_x \prod_{c \in C} \phi_c(\mathbf{z}, \mathbf{x_c})$, which sums over all possible label assignments. In the log-linear parameterisation of CRFs, the potentials are described using a log linear combination of feature functions:

$$\phi_c(\mathbf{z}, \mathbf{x_c}) = \exp\left(\mathbf{w}_c^T \cdot \mathbf{f_c}(\mathbf{z}, \mathbf{x_c})\right) \tag{2}$$

where $\mathbf{w}_c^T$ is the weight vector we need to learn and $\mathbf{f_c}(\mathbf{z}, \mathbf{x_c})$ is a feature function extracting a vector of features in a clique. Here, large weights imply high compatibility and low weights a less likely configuration. Thus, the log-linear parameterisation of a CRF can be written as:

$$p(\mathbf{x}|\mathbf{z}) = \frac{1}{Z(\mathbf{z})} \exp\left\{\sum_{c \in C} \mathbf{w}_c^T \cdot \mathbf{f_c}(\mathbf{z}, \mathbf{x_c})\right\} \tag{3}$$

where $Z(\mathbf{z}) = \sum_x \exp\left\{\sum_{c \in C} \mathbf{w}_c^T \cdot \mathbf{f_c}(\mathbf{z}, \mathbf{x_c})\right\}$ is the new partition function.

CRFs avoid encoding distributions over just the variable $\mathbf{z}$ which allows great flexibility in choosing a rich set of continuous observed variables (feature functions) whose dependencies may be difficult to model (no simple parametric form) or understand. We can therefore exploit domain knowledge to define a set of discriminative features without needing to model their joint distribution [29].

C. Conditional Log Likelihood Training and Inference

In parameter estimation, we are given the training dataset of M examples and associated graphical structures. Our aim is to estimate the parameters $\mathbf{w}_c^T$ in Equation (3) that best model the data. Let $\theta = [\mathbf{v}, \mathbf{w}]$ represent the combined parameters of the unary and binary features and $F(\mathbf{z}, \mathbf{x})$ represent all the features. The CRF learns the weights discriminatively by maximising the conditional log likelihood of the training data. Specifically, the log likelihood for M training examples is given by:

$$\log(p(\mathbf{x}|\mathbf{z})) = \sum_{m=1}^{M} \theta^T F(\mathbf{z}_m, \mathbf{x}_m) - \sum_{m=1}^{M} \log \mathbf{Z}(\theta, \mathbf{z}_m) \tag{4}$$

This log likelihood function can be shown to be concave because $\log \mathbf{Z}(\theta, \mathbf{z})$ is concave [29]. No analytical form for this log likelihood maximum exists thus we perform parameter estimation using an L2-regularised log-likelihood and estimate the maximum using an iterative line search method (quasi-Newton approach) which takes steps in parameter space to improve the objective. L2-regularisation is necessary because as more features are added to the CRF model, the negative log-likelihood function decreases which could lead to over-fitting.

The regularisation essentially adds a penalty for large parameter values [29].

Inference refers to estimating the marginal distribution of each hidden variable $\mathbf{x_i}$ or the most likely configuration of all hidden variables $\mathbf{x}$. These tasks rely on using belief propagation, a message passing algorithm for estimating marginal distributions over a graph [30]. In belief propagation, messages are passed through the graph structure of the model and each node passes on received messages to its neighbours. Although this method is exact for trees, forests and other graphs not containing loops, our Delaunay triangulation connectivity creates loopy graphs, therefore we apply loopy belief propagation (LBP) [31], an approximate inference algorithm not guaranteed to necessarily converge to the true probability distribution (inference is intractable). We however found the performance of the LBP algorithm for inference in our CRF model to produce good results, even though convergence is not necessarily guaranteed.

D. Feature Functions

Feature functions for road marking classification describe geometric and spatial properties of line segments and their endpoints. The CRF learning algorithm associates an appropriate weight with each feature function given the training data. In CRFs, local (unary) features depend only on a single hidden state i. Global (binary) features depend on two hidden states i and j. In this section we discuss the visual feature functions, unary score features, and binary features used.

1) Visual Feature Functions: We formulate eleven geometric and spatial features capturing the properties of a single node. Let us denote the line segment endpoints for node i as p_{Ai} and p_{Bi}, the midpoint as p_{Ci}, and the x and y segment endpoints for node i as ($p_{Ai,x}$ and $p_{Bi,x}$) and ($p_{Ai,y}$ and $p_{Bi,y}$), respectively. Let p_{Ck} be the midpoint of the node closest to node i. Let $isintersect()$ be a function that returns 1 if there is an intersection and 0 otherwise, and l_i denote the line segment formed by joining p_{Ai} and p_{Bi}. Then the visual feature functions are:

Length: This feature measures the length of a line segment and is useful for separating typically long road markings (boundary, separator), medium length (zig-zag, special lane) and short markings (intersection and special lane).

$$f_{length}(i, \mathbf{z}_i) = \| p_{Ai} - p_{Bi} \| \tag{5}$$

X-midpoint location: This feature measures the spatial x location of the road marking in the image. This captures the fact that boundary or zig-zag type markings tend to be located toward the edges of the image while boxed junctions tend to take up the entire image. Similarly, special lanes and intersections are spatially located near the road boundaries.

$$f_{midx}(i, \mathbf{z}_i) = p_{Ci,x} \tag{6}$$

Y-midpoint location: For completion we also consider the y-midpoint location of the road marking. Although we removed the perspective distortion in the image, this feature is useful

particularly at intersections and roundabouts where the overall orientation of road markings appears rotated.

$$f_{midy}(i, \mathbf{z}_i) = p_{Ci,y} \tag{7}$$

Delta-x: In addition to capturing the location and length of a road marking line segment, it is also important to have features that capture the regularities. The change in x coordinates (delta x) feature captures the regular pattern of zig-zag approaches. We showed that markings of the same length and one of two orientations denote zig-zags. The delta x (and delta y) feature is constant for zig-zag type markings. For vertical type markings (boundaries, special lanes etc.) this feature is close to zero.

$$f_{deltax}(i, \mathbf{z}_i) = \| p_{Ai,x} - p_{Bi,x} \| \tag{8}$$

Delta-y: This feature captures similar geometric properties to Delta-x. Specifically, for zig-zag markings, this feature is constant. For vertical markings it simply reduces to the approximate length of the road marking and for boxed junctions this feature varies depending on the type of boxed junction.

$$f_{deltay}(i, \mathbf{z}_i) = \| p_{Ai,y} - p_{Bi,y} \| \tag{9}$$

Orientation: The orientation of a line segment is measured anticlockwise relative to the x-axis. This feature approximately evaluates to one of two values for zig-zag type markings and, with the exception of the boxed junction class, is approximately constant for road markings belonging to the same class in an image. It also roughly captures the vehicle orientation (which influences the relative orientation of the road markings).

$$f_{angle}(i, \mathbf{z}_i) = \angle(p_{Ai} - p_{Bi}, \text{x} - \text{axis}) \tag{10}$$

Furthest vertical distance: This feature is simply a measure of how far away from the top of the image a road marking is. Although this is sensitive to the vehicle pitch, for a given vehicle class, the relative vertical distances across all road markings in the image is preserved.

$$f_{vert}(i, \mathbf{z}_i) = \max(p_{Ai,y}, p_{Bi,y}) \tag{11}$$

Proportion segments to the left: This feature measures the percentage of road markings located to the left of the ith node. This captures the overall distribution of road markings in the image relative to the position of node i. Intuitively, we expect this class distribution to be different across classes, with boundary type markings having low or high proportions and separator markings proportions close to 0.5. Let N be the total number of nodes in an image and K_{left} be the number of nodes to the left of node i.

$$f_{propleft}(i, \mathbf{z}_i) = \frac{K_{left}}{(N-1)} \tag{12}$$

Proportion segments to the right: This feature measures the percentage of road markings located to the right of the ith node. It captures the overall distribution of road markings in the image relative to the position of node i similar to the previous feature. The symmetry of the road markings on the road is also

TABLE II
CLASS DISTRIBUTION OF OUR DATASET SHOWING THE NUMBER
AND PROPORTION OF TRAINING AND TEST EXAMPLES

Class	Description	Train	Proportion	Test	Proportion
1	Single Boundary	794	12.19%	778	11.81%
2	Double Boundary	1012	15.54%	961	14.59%
3	Separator	1491	22.89%	1471	22.34%
4	Zig-zag	523	8.03%	549	8.34%
5	Intersection	287	4.41%	395	6.00%
6	Boxed Junction	1566	24.04%	1573	23.89%
7	Special Lane	841	12.91%	858	13.03%

captured by this and the previous proportion feature. Let N be the total number of nodes in an image and K_{right} be the number of nodes to the right of node i.

$$f_{propright}(i, \mathbf{z}_i) = \frac{K_{right}}{(N-1)} \quad (13)$$

Closest midpoint distance: This feature is especially important for double boundary type road markings since they comprise of two road markings of similar length and orientation, running parallel. Thus the midpoint distance between double boundary road markings is small compared to the midpoint distances between other marking types.

$$f_{midist}(i, k, \mathbf{z}_i) = \|p_{Ci} - p_{Ck}\| \quad (14)$$

Number of intersecting line segments: This feature is most important for boxed junction markings occurring at intersections/junctions. This is because for other classes, mostly no intersections occur. The feature function counts the number of line segments k that intersect with the line segment associated with node i.

$$f_{int}(i, k, \mathbf{z}_i) = \sum_{k=1}^{N-1} isintersect(l_i, l_k) \quad (15)$$

Conditional Random Fields can incorporate continuous local features like the ones described above. CRF potentials are formed by exponentiating a linear function of the feature functions. To model the more complex nonlinear relationships between hidden states and feature values however, it is better to discretise (or binarise) the features [32].

2) Unary Score Features: Instead of directly using the visual features described in Section IV-D1 in our CRF model, we first train a Random Under Sampling Boost (RUSBoost) [33] classifier and use these scores as input to our CRF. This is a design choice shown to be effective by Friedman *et al.* [34] who proposed first learning AdaBoost classifiers from trained labelled data. Other variants of this approach, include using SVM scores and CRFs [32] to build a two stage classifier. Specifically, we choose RUSBoost because it is robust to skewed training data—the number of examples, per class, in our datasets is not balanced. For example, separator markings occur over five times more frequently than intersection type markings (see Table II).

We train a RUSBoost classifier using the eleven features described in Section IV-D1. The classification output for each training example is a 7×1 vector assigning a score to each class. The RUSBoost classifier learns an independent model

which does not account for relationships between adjacent nodes, thus we use this score vector as input features for our CRF which learns statistical dependencies between adjacent hidden states.

3) Binary Features: For consistency (relational) across the graphical model, it is necessary to define features relating hidden states in the CRF to each other. Such pairwise features take into account the effect that the feature functions have on the transitions from one state to another (for example a neighbour node transitioning from a zig-zag to a boxed junction class). We derive edge (binary) features from the RUSBoost scores (new node features). The edge features are taken to be the node features from both edge ends [35].

V. EXPERIMENTAL RESULTS

We created a benchmarking dataset containing 2068 hand labelled images spanning city, residential and motorway roads [7]. The road marking dataset class distribution is shown in Table II where a skewed distribution is noted. Some classes, such as intersections occur less frequently than others such as separators.

We evaluate our system performance in three ways:

- Without relational learning (similar to Duchow and Kortner [21])
- With relational learning (our new proposed method)
- Sensitivity to input errors

Our road marking classification framework was implemented using the Undirected Graphical Models (UGM) Toolbox by Mark Schmidt [35] for probabilistic graphical models.

A. Results

1) No Relational Learning: The performance of our road marking classification framework without relational learning (only RUSBoost classifier) was evaluated using the dataset described above. Although the three feature functions, $f_{propleft}$, $f_{propright}$, and f_{midist} take other road markings into account, the relationships between adjacent node labels (or cliques) is not explicitly modeled. The resulting per class precision and recall confusion matrices are shown in Table III. For a given confusion matrix, the diagonal elements are large ratios (close to 1) for good classification performance (i.e. the ground truth and predicted classes correspond exactly). Thus, a strong bright diagonal in a confusion matrix illustrates good classification performance.

Our experiments resulted in per class precision values ranging between 0.56 and 0.89 and per class recall ranging from 0.54 to 0.96 as shown in Table III. The intersection class precision is very low because examples belonging to this class are commonly misclassified as special lanes, boxed junctions or sometimes zig-zag. Per class recall values for the special lane class are also very low because road markings predicted as separator, single boundary, zig-zag approach, intersection or

TABLE III
PER CLASS PRECISION AND RECALL CONFUSION MATRICES WITHOUT
RELATIONAL LEARNING. (a) PER CLASS PRECISION MATRIX
(NO RELATIONAL LEARNING). (b) PER CLASS RECALL
MATRIX (NO RELATIONAL LEARNING)

PREDICTION

	SGL	DBL	DIV	ZIG	INT	BJ	SPL
SGL	**0.878**	0.008	0.024	0.004	0.001	0.033	0.052
DBL	0.007	**0.85**	0.007	0.007	0.007	0.096	0.026
DIV	0.01	0	**0.867**	0.003	0.003	0.033	0.085
ZIG	0.003	0.007	0.01	**0.712**	0.053	0.154	0.061
INT	0	0.016	0.004	0.044	**0.558**	0.193	0.185
BJ	0.003	0.01	0.024	0.012	0.019	**0.89**	0.042
SPL	0.009	0.021	0.051	0.015	0.066	0.124	**0.713**

(GROUND TRUTH)

(a)

PREDICTION

	SGL	DBL	DIV	ZIG	INT	BJ	SPL
SGL	**0.955**	0.007	0.014	0.005	0.003	0.018	0.051
DBL	0.009	**0.952**	0.005	0.015	0.02	0.065	0.033
DIV	0.021	0	**0.933**	0.007	0.01	0.034	0.156
ZIG	0.003	0.005	0.005	**0.887**	0.091	0.067	0.049
INT	0	0.008	0.001	0.04	**0.709**	0.062	0.108
BJ	0.005	0.012	0.02	0.027	0.058	**0.703**	0.061
SPL	0.008	0.015	0.022	0.018	0.109	0.051	**0.542**

(GROUND TRUTH)

(b)

Fig. 6.   Special lane and intersection type road markings have very similar visual appearance (short, repetitive).

Fig. 7.   Boxed junctions describe areas of the road that a vehicle should not be stationary in (but may possibly quickly traverse) such as separator boxed junctions (left) and intersection boxed junctions (right).

boxed junction actually belong to the special lane class. The RUSBoost model is discriminative and assigns road marking classes based on a strong learner given the feature functions. We show in the next section how introducing a relational learning

TABLE IV
PERCENTAGE IMPROVEMENTS IN TERMS OF PRECISION AND
RECALL WHEN RELATIONAL LEARNING IS USED

Class	Name	Precision Change (%)	Recall Change (%)
1	SGL	5.4	-1.4
2	DBL	5	-2.4
3	DIV	1.2	-1.8
4	ZIG	18.6	4.2
5	INT	23.4	-2.3
6	BJ	1.6	17.6
7	SPL	2.3	17.9

TABLE V
PER CLASS PRECISION AND RECALL CONFUSION MATRICES WITH
RELATIONAL LEARNING. (a) PER CLASS PRECISION MATRIX
(WITH RELATIONAL LEARNING). (b) PER CLASS RECALL
MATRIX (WITH RELATIONAL LEARNING)

PREDICTION

	SGL	DBL	DIV	ZIG	INT	BJ	SPL
SGL	**0.932**	0.008	0.018	0	0.001	0.013	0.028
DBL	0.003	**0.9**	0.01	0.002	0.021	0.044	0.019
DIV	0.014	0.002	**0.879**	0.005	0.003	0.032	0.065
ZIG	0.002	0.004	0.005	**0.898**	0.051	0.032	0.009
INT	0	0.02	0.009	0.003	**0.792**	0.041	0.135
BJ	0.003	0.022	0.017	0.007	0.015	**0.906**	0.031
SPL	0.019	0.02	0.082	0.021	0.055	0.067	**0.736**

(GROUND TRUTH)

(a)

PREDICTION

	SGL	DBL	DIV	ZIG	INT	BJ	SPL
SGL	**0.941**	0.006	0.01	0	0.003	0.006	0.026
DBL	0.004	**0.928**	0.007	0.004	0.053	0.028	0.022
DIV	0.028	0.003	**0.915**	0.015	0.01	0.031	0.117
ZIG	0.001	0.002	0.002	**0.929**	0.073	0.011	0.006
INT	0	0.007	0.002	0.002	**0.686**	0.009	0.054
BJ	0.005	0.035	0.018	0.018	0.058	**0.879**	0.055
SPL	0.021	0.018	0.047	0.033	0.116	0.036	**0.721**

(GROUND TRUTH)

(b)

framework, which models how adjacent markings are likely to belong to the same class, reduces the misclassification errors.

2) Classification With Relational Learning: We further tested our system using our CRF relational learning framework. Here, relationships between adjacent road marking classes are explicitly captured. The posterior label probabilities produced by the CRF encode our uncertainty in the predictions, a property which is useful for decision making.

The resulting per class precision and recall confusion matrices are shown in Table V—the per class precision values range between 0.74 and 0.93 while the per class recall ranges from 0.69 to 0.94. The precision is lowest for the intersection and special lane classes as their visual appearance is similar (see Fig. 6) and the intersection class is under represented in the training data (4.41% of the training data). Similarly, the recall is lowest for the intersection and special lane classes. The remaining classes show very good performance with precision and recall values of between 0.88 to 0.94. We quantify the per class percentage improvement gained by adding relational information in the road marking classification framework in Table IV, where performance improvements of up to 23% are achieved
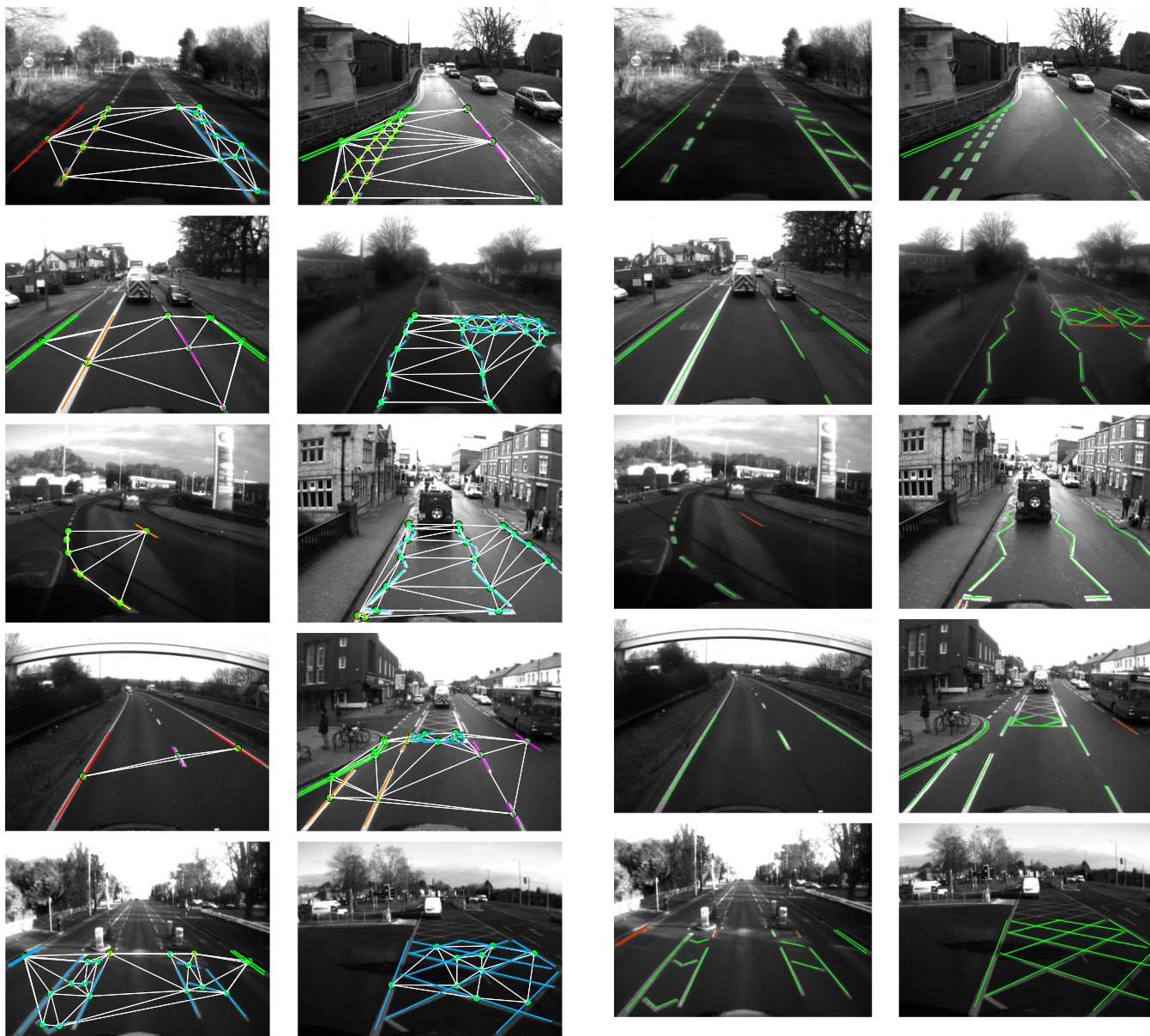
Fig. 8. Classification results where each colour represents a different class (see Table I for colour code). The graphical model structure is superimposed on the image.

Fig. 9. Classification results from our system (green lines are correctly classified, whilst red lines represent misclassifications). These images correspond to Fig. 8.

across classes. Table IV shows that relational learning results in slightly reduced recall performance for single boundary, double boundary, separator, and intersection classes. This is because there is not much relational structure in these types of road markings. The remaining classes however contain relational structure thus precision and recall performance is significantly increased when relational learning is used (see Table IV). Our classifier thus also learns which road marking classes are relational.

Fig. 8 shows classification results where each colour represents the class the road marking is assigned to. To better visualise the errors, we plot these same images in Fig. 9 where green represents correctly classified road markings, and red lines show the misclassifications.

Our experiments show that the amount of per class training data affects the classification accuracy of our system. In particular, the intersection class (which has the worst performance) is also the most under represented in the training set. However, because CRF features do not necessarily completely specify a state/observation, the model can be estimated from less training data [27]. Although no analytical method exists for determining the amount of data required to sufficiently train the algorithm, we found the dataset we used to be sufficient to identify all the hidden nodes.

3) Sensitivity to Noise: To test the sensitivity of our system to input errors (poor detection of road markings) we simulated noise resulting from vehicle occlusions, shadows or faded road markings. For a given road scene, the system performance (misclassification error rate) was calculated for varying degrees

Fig. 11. Some zig-zag approaches correctly classified by our algorithm, with 88% accuracy. Thus our vehicle can anticipate upcoming traffic light controlled pedestrian crossings.
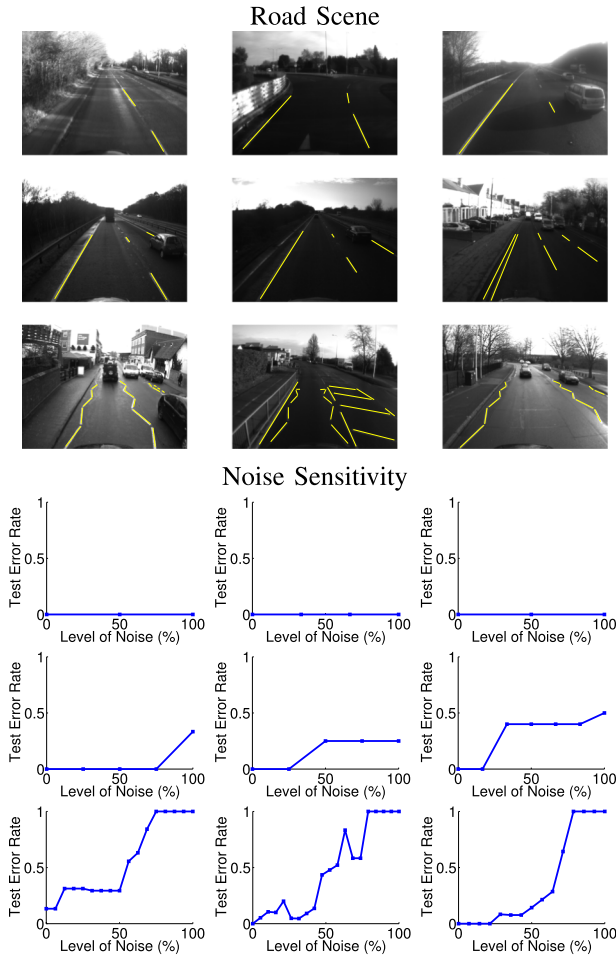
Fig. 10. The sensitivity of the road marking classifier to imperfect discrimination of road markings. For multi-lane road scenes (top and middle row), the error rate remains below 0.5 for noise levels of up to 83%, thus showing a robust system. Semantically rich road scenes (bottom row) are however more dramatically affected by input noise.

Fig. 12. Examples of boxed junctions correctly classified by our algorithm (86% prediction accuracy). Thus no stopping or traversal areas are automatically detected.

of input noise as shown in Fig. 10. Multi-lane road scenes were found to be more robust to noise compared to more semantically rich road scenes.

VI. CONCLUSION AND FUTURE WORK

In this paper, we presented a framework for situational awareness based on reading and interpreting road markings. Specifically, we classified road markings into seven distinct classes: single boundary, double boundary, separator, zig-zag, intersection, boxed junction and special lane. Our system achieved a precision of between 74% and 93% and corresponding recall of 69% to 94% across all classes. Overall, our system allows a vehicle to read and interpret the road by understanding the rules of the road encoded in the type of painted road markings visible.

In future work, we will investigate using the output of our road marking classification system for road scene semantic interpretation for situational awareness. To test this, we performed some preliminary experiments to detect traffic light controlled pedestrian crossings (zig-zag approaches) and no drive regions (boxed junctions—see Fig. 7).

Here, we defined a zig-zag approach (or boxed junction) to exist if a proportion of over 0.7 of classified individual road marking segments belong to class zig-zag (or boxed junction). This threshold was determined empirically to capture our observation that zig-zag approaches (or boxed junction) occur when the vast majority of road markings in a road scene belong to that class. Our experimental results showed that, even with this straight forward approach, our classifier achieved 88% and 86% prediction accuracies for the zig-zag approach and boxed junction scenes, respectively. Examples of correctly classified scenes are shown in Figs. 11 and 12, respectively. A vehicle can therefore prepare for the possibility of encountering humans on the road, or areas where it may not be stationary—thus gaining a dynamic sense of situational awareness, like a human. Investigating this further forms the subject of our future work.

References

[1] M. Endsley, "Towards a theory of situation awareness in dynamic systems," *Human Factors*, vol. 37, no. 1, pp. 32–64, Mar. 1995.

[2] A. Wimmer, T. Jungel, M. Glück, and K. Dietmayer, "Automatic generation of a highly accurate map for driver assistance systems in road construction sites," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2010, pp. 281–286.

[3] N. Fairfield and C. Urmson, "Traffic light mapping and detection," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 5421–5426.

[4] B. Mathibela, M. Osborne, I. Posner, and P. Newman, "Can priors be trusted? Learning to anticipate roadworks," in *Proc. IEEE Conf. Intell. Transp. Syst.*, Anchorage, AK, USA, 2012, pp. 927–932.

[5] B. Mathibela, I. Posner, and P. Newman, "A roadwork scene signature based on the opponent color model," in *Proc. IEEE/RSJ Int. Conf. IROS*, 2013, pp. 4394–4400.

[6] G. Emmerson, "Minimising disruption from roadworks," Transport for London Surface Transp. Panel, London, U.K., May 2010.

[7] B. Mathibela, "Situational awareness in autonomous vehicles: Learning to read the road," Ph.D. dissertation, Univ. Oxford, Oxford, U.K., 2014.

[8] T. Veit, J.-P. Tarel, P. Nicolle, and P. Charbonnier, "Evaluation of road marking feature extraction," in *Proc. Int. IEEE Conf. Intell. Transp. Syst.*, 2008, pp. 174–181.

[9] T. Wu and A. Ranganathan, "A practical system for road marking detection and recognition," in *Proc. IEEE Intell. Veh. Symp.*, 2012, pp. 25–30.

[10] P. Charbonnier, F. Diebolt, Y. Guillard, and F. Peyret, "Road markings recognition using image processing," in *Proc. IEEE Conf. Intell. Transp. Syst.*, 1997, pp. 912–917.

[11] P. Foucher, Y. Sebsadji, J.-P. Tarel, P. Charbonnier, and P. Nicolle, "Detection and recognition of urban road markings using images," in *Proc. 14th Int. IEEE Conf. Intell. Transp. Syst.*, 2011, pp. 1747–1752.

[12] S. Vacek, C. Schimmel, and R. Dillmann, "Road-marking analysis for autonomous vehicle guidance," in *Proc. Eur. Conf. Mobile Robots*, 2007, pp. 1–6.

[13] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 1, pp. 886–893.

[14] J. Collado, C. Hilario, A. de la Escalera, and J. Armingol, "Detection and classification of road lanes with a frequency analysis," in *Proc. IEEE Intell. Veh. Symp.*, 2005, pp. 78–83.

[15] J. Collado, C. Hilario, A. D. L. Escalera, and J. Armingol, Adaptative road lanes detection and classification, *Advanced Concepts for Intelligent Vision Systems*. New York, NY, USA: Springer-Verlag, 2006, pp. 1151–1162.

[16] B. Soheilian, N. Paparoditis, and D. Boldo, "3D road marking reconstruction from street-level calibrated stereo pairs," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 4, pp. 347–359, Jul. 2010.

[17] R. Danescu and S. Nedevschi, "Detection and classification of painted road objects for intersection assistance applications," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst.*, 2010, pp. 433–438.

[18] L. Yunchong, H. Kezhong, and J. Peifa, "Road markers recognition based on shape information," in *Proc. Intell. Veh. Symp.*, 2007, pp. 117–122.

[19] A. Kheyrollahi and T. Breckon, "Automatic real-time road marking recognition using a feature driven approach," *Mach. Vis. Appl.*, vol. 23, no. 1, pp. 123–133, Jan. 2012.

[20] M. Noda *et al.*, "Recognition of road markings from in-vehicle camera images by a generative learning method," in *Proc. Mach. Vis. Appl.*, 2009, pp. 514–517.

[21] C. Duchow and B. Kortner, "Aggregating lane markings into lanes for intersection assistance," in *Proc. IEEE Intell. Veh. Symp.*, 2007, pp. 722–727.

[22] G. Y. Jiang, T. Y. Choi, S. K. Hong, J. W. Bae, and B. S. Song, "Lane and obstacle detection based on fast inverse perspective mapping algorithm," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2000, vol. 4, pp. 2969–2974.

[23] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, vol. 2. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[24] H. Ishida, K. Kidono, Y. Kojima, and T. Naito, "Road marking recognition for map generation using sparse tensor voting," in *Proc. 21st Int. Conf. Pattern Recognit.*, 2012, pp. 1132–1135.

[25] A. Huang, D. Moore, M. Antone, E. Olson, and S. Teller, "Finding multiple lanes in urban road networks with vision and Lidar," *Autom. Robots*, vol. 26, no. 2/3, pp. 103–122, 2009.

[26] D.-T. Lee and B. Schachter, "Two algorithms for constructing a Delaunay triangulation," *Int. J. Comput. Inf. Sci.*, vol. 9, no. 3, pp. 219–242, Jun. 1980.

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

[28] F. Ramos, D. Fox, and H. Durrant-Whyte, "CRF-Matching: Conditional random fields for feature-based scan matching," in *Proc. Robot. Sci. Syst.*, 2007, pp. 1–8.

[29] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT Press, 2009.

[30] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.

[31] K. Murphy, Y. Weiss, and M. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proc. Conf. Uncertainty Artif. Intell.*, Morgan Kaufmann, San Mateo, CA, USA, 1999, pp. 467–475.

[32] G. Hoefel and C. Elkan, "Learning a two-stage SVM/CRF sequence classifier," in *Proc. ACM Conf. Inf. Knowl. Manag.*, 2008, pp. 271–278.

[33] C. Seiffert, T. Khoshgoftaar, J. van Hulse, and A. Napolitano, "RUSBoost: Improving classification performance when training data is skewed," in *Proc. 19th Int. Conf. Pattern Recognit.*, 2008, pp. 1–4.

[34] S. Friedman, H. Pasula, and D. Fox, "Voronoi random fields: Extracting topological structure of indoor environments via place labeling," in *Proc. Int. Joint Conf. Artif. Intell.*, 2007, vol. 7, pp. 2109–2114.

[35] M. Schmidt, (2010) UGM: A Matlab Toolbox for Probabilistic Undirected Graphical Models. [Online]. Available: http://www.di.ens.fr/mschmidt/Software/UGM.html

Bonolo Mathibela received the B.Sc. degree in mechatronics engineering from University of Cape Town, Cape Town, South Africa, and the D.Phil. degree in engineering science from University of Oxford, Oxford, U.K.

She was with the Mobile Robotics Group, Department of Engineering Science, University of Oxford. She tackles the question of road semantic understanding and interpretation for self-driving cars and intelligent vehicles. Her research interests include situational awareness in intelligent vehicles, specifically biologically inspired roadwork scene understanding, colour spaces, roadwork mapping, road marking detection and interpretation, road boundary estimation, and semantically driven vehicle localisation.

Paul Newman received the M.Eng. degree in engineering science from University of Oxford, Oxford, U.K., and the Ph.D. degree in autonomous navigation from University of Sydney, Sydney, Australia.

In 1999, he worked in the commercial subsea navigation industry. In late 2000, he joined the Department of Ocean Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA, as a Postdoctoral Researcher, later becoming a Research Scientist. In early 2003, he returned to University of Oxford as a Departmental Lecturer in engineering science before becoming a Fellow of New College in 2005, a Professor of engineering science in 2010, and a BP Professor of information engineering and a Fellow of Keble College in 2012. He is currently with the Mobile Robotics Group, Department of Engineering Science, University of Oxford.

Ingmar Posner received the M.Eng. degree in electronic systems engineering from Aston University, Birmingham, U.K., and the D.Phil. degree in bioacoustics from University of Oxford. He is a University Lecturer with University of Oxford, where he coleads the Mobile Robotics Group, Department of Engineering Science. His research tackles semantic information inference from the environment traversed and how this knowledge can feed into the decision-making process of an autonomous agent. His research focus includes closing the action–perception loop in semantic mapping to enable robust robot decision making, online planning, and exploration.