Information Sciences 000 (2017) 1–16

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Hybrid conditional random field based camera-LIDAR fusion for road detection

Liang Xiao[a], Ruili Wang[b], Bin Dai[a,*], Yuqiang Fang[a], Daxue Liu[a], Tao Wu[a]

[a] College of Mechatronic Engineering and Automation, National University of Defense Technology, Changsha, Hunan, China
[b] Institute of Natural and Mathematical Sciences, Massey University, Auckland, New Zealand

ARTICLE INFO

Article history:
Received 13 December 2016
Revised 23 April 2017
Accepted 28 April 2017
Available online xxx

Keywords:
Sensor fusion
Road detection
Condition random field
Boosted decision tree

ABSTRACT

Road detection is one of the key challenges for autonomous vehicles. Two kinds of sensors are commonly used for road detection: cameras and LIDARs. However, each of them suffers from some inherent drawbacks. Thus, sensor fusion is commonly used to combine the merits of these two kinds of sensors. Nevertheless, current sensor fusion methods are dominated by either cameras or LIDARs rather than making the best of both. In this paper, we extend the conditional random field (CRF) model and propose a novel hybrid CRF model to fuse the information from camera and LIDAR. After aligning the LIDAR points and pixels, we take the labels (either road or background) of the pixels and LIDAR points as random variables and infer the labels via minimization of a hybrid energy function. Boosted decision tree classifiers are learned to predict the unary potentials of both the pixels and LIDAR points. The pairwise potentials in the hybrid model encode (i) the contextual consistency in the image, (ii) the contextual consistency in the point cloud, and (iii) the cross-modal consistency between the aligned pixels and LIDAR points. This model integrates the information from the two sensors in a probabilistic way and makes good use of both sensors. The hybrid CRF model can be optimized efficiently with graph cuts to get road areas. Extensive experiments have been conducted on the KITTI-ROAD benchmark dataset and the experimental results show that the proposed method outperforms the current methods.

1. Introduction

Road detection is a fundamental research topic in autonomous vehicles and has been studied for decades [11]. For autonomous vehicles, stable and accurate road detection is a prerequisite. As there are different kinds of roads, such as highways, urban roads and country roads, with different features, the approaches to detect them are different. In well painted highways, road detection can be replaced by lane detection, which is considered to be much easier. However, it is much more challenging to detect normal urban roads for many reasons such as the variations in road materials from segment to segment, the similarities of textures and heights between the road areas and non-road areas, the changes of illumination and weather and so on.

* Corresponding author.
E-mail addresses: xiaoliang@nudt.edu.cn (L. Xiao), r.wang@massey.ac.nz (R. Wang), bindai.cs@gmail.com (B. Dai), fangyuqiang@nudt.edu.cn (Y. Fang), liudaxue78@gmail.com (D. Liu), wutao@nudt.edu.cn (T. Wu).

http://dx.doi.org/10.1016/j.ins.2017.04.048
0020-0255/

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048

To achieve accurate and stable road detection, many algorithms based on different kinds of sensors have been developed. The most commonly used sensors are monocular cameras [4,15] and LIght Detection And Rangings (LIDARs) [8] which can acquire different kinds of information for road detection. Monocular vision captures the perspective projection of the scene and then the dense colors and textures can be used to group the pixels or super-pixels into road and background areas. However, monocular vision often suffers from the changes of illumination and weather, and it cannot capture accurate 3D information. Compared to vision, a LIDAR is an active sensor which works independently of the ambient light and it can measure the distances to objects accurately. However, in the point clouds captured by a LIDAR, neither color nor texture information is available and the points are rather sparse.

To overcome the inherent drawbacks and combine the merits of different kinds of sensors, multi-modal sensor fusion has been widely used [31,33,43,50,53]. For road detection, several camera-LIDAR fusion methods have been proposed. However, most of them are dominated by either cameras or LIDARs and fail to fully exploit the advantages of both sensors. For example, in [43], after projecting the LIDAR point cloud onto the image, the feature used for obstacle classification is dominated by the height information of the LIDAR points while the pixel information is ignored. In [21], the information from the images and LIDAR point clouds is utilized separately in a stage-wise fashion. The LIDAR point clouds are only used for ground seed extraction, while the following road detection and segmentation are dominated by the image. In [54], fusion is performed on the feature and region levels, resulting in a coarse level fusion. All these methods fail to fuse the image and LIDAR in fine granularity and through a joint model. This work aims to fill this gap. Aside from the multi-modal information, another kind of information that is crucial for improving the performance is the contextual information in each modality. Considering the strength of conditional random fields (CRF) in modeling contextual information [44], we extend CRF to a multi-modal setting and propose a novel hybrid-CRF-based camera-LIDAR fusion method to improve the performance of road detection. By formulating the road detection as a binary labeling problem, the labels (either road or background) of the pixels and LIDAR points are taken as random variables and a hybrid CRF model is built to solve the multi-modal labeling problem. The proposed method utilizes the learned boosted decision tree classifiers to derive the unary potentials of the pixels and LIDAR points. The neighboring smooth prior of the pixels and LIDAR points, together with the consistency constraint between the aligned LIDAR points and pixels are modeled via the pairwise potentials. This model integrates the information from the two sensors probabilistically and the information from both sensors are well exploited. The hybrid CRF model can be optimized efficiently by graph cuts [23] to get road areas. Experiments conducted on the KITTI-ROAD benchmark dataset [14] demonstrate that the proposed hybrid CRF model is effective in fusing multi-modal information and the results of road detection are better compared to that of the current existing methods.

The main contributions of this paper include: (i) A novel hybrid CRF model is proposed to fuse the image and LIDAR point cloud, in which the contextual consistency of the image and LIDAR point cloud, together with the constraint of cross-modal consistency is jointly modeled probabilistically, and (ii) the proposed sensor fusion framework is applied to urban road detection and our method achieves good performance on the KITTI-ROAD benchmark dataset [14]. The results of our method on the UM subset rank first on the leaderboard [1] apart from the deep-learning-based ones, which usually rely on models pre-trained on extra data for initialization and modern GPUs for fast computing.

The rest of this paper is organized as follows. Section 2 reviews the work on road detection. Section 3 shows how the LIDAR points and the images are registered. In Section 4, we first introduce the CRF-based labeling framework, then we provide the detailed information about the proposed hybrid CRF model. The training of the pixel and LIDAR point classifiers is described in Section 5, along with the feature extraction. The experimental results tested on the KITTI-ROAD benchmark dataset are given in Section 6. Finally, conclusions and directions for the future work are listed in Section 7.

2. Related work

As a fundamental problem in developing autonomous vehicles, road detection has been extensively studied. Various road detection systems have been developed based on different kinds of sensors as well as fusion of some sensor types.

The most frequently used sensor for road detection is the monocular camera [20]. Monocular-vision-based road detection is usually formulated as a classification problem, i.e., classifying each pixel or super-pixel into either road or background. Many kinds of machine learning methods have been applied to road detection, such as mixture of Gaussian [10], support vector machines [2], extreme learning machines [30,55], neural networks [42], boosting [15] and structured random forest [51]. In recent years, many new feature learning methods have been applied to road detection such as slow feature analysis [15], sparse coding and dictionary learning [28,29,32,52], convolutional neural network [3,35] and deep deconvolutional network [37]. Classification-based methods classify each unit independently and do not take the contextual interaction into consideration. Therefore, the prediction may be noisy. To solve this problem, conditional random fields (CRF) [19,41,45,50] are widely used to model the contextual interaction. Generally, CRF-based methods are supposed to get better performance than simple classification-based methods. However, when the image quality is badly affected by illumination or weather conditions, these methods may also get poor results.

LIDAR is another widely used sensor in autonomous vehicles. Various LIDAR-based road detection algorithms have been proposed and they can be roughly categorized into two groups: regression-based and classification-based algorithms. Based on the assumption of the continuity of the road area, the regression-based algorithms utilize one dimensional curve fitting [8,21] or two dimensional surface fitting [5,12] to segment the road. The classification-based algorithms extract features of the points or grid cells and then classify them based on certain intuitive rules or learning methods, such as elevation map

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048

analysis [46], Gaussian Mixture Model [26] and local convexity criterion [38]. Similar to image-based algorithms, Markov random fields (MRF) can be employed to model the contextual information of the LIDAR points to get locally consistent results. The random fields can be built on the grid map [17], the cylindrical grid map [7] or the neighboring graph of points [40]. In a nutshell, LIDAR-based road detection algorithms analyze the 3D information of the point cloud to get the obstacle-free area as the road. However, in some scenes, the roadside areas have no significant differences in heights with the road areas and then these methods may fail.

Since camera and LIDAR both have some drawbacks, sensor fusion becomes a natural solution to overcome the inherent defects of each single sensing modality. Recently, camera-LIDAR fusion has been applied to road detection. Shinzato et al. [43] proposed a simple and efficient sensor fusion method to detect the road terrain. This method firstly projects the LIDAR points onto the image plane and constructs a graph by Delaunay triangulation. Then, the nodes are classified into obstacles and non-obstacles. Finally, multiple free space detections are employed to get the dense road area in the image plane. However, this method did not actually utilize any pixel information. It only used the cross calibration parameters to get the LIDAR points projected onto the image plane. Hu et al. [21] proposed a more intuitive method to fuse the information from LIDAR and camera. Plane estimation was employed to extract the ground points in a LIDAR point cloud. These points were projected onto the image for learning a Gaussian model of the illumination invariant image feature by which the pixels were classified. This method used the LIDAR points to generate the seeds for image-based segmentation in a stage-wise fashion. In other words, in the first stage, only the LIDAR point cloud was used to extract the seed ground points. Then, in the second stage, the image was segmented according to the model learned from the seed pixels, while the LIDAR point cloud was totally discarded. Compared to the stage-wise method, we argue that joint modeling the information of both camera and LIDAR at the same time via the CRF framework will be more beneficial.

Although CRF has been widely used in image labeling and LIDAR point cloud labeling, the ability of CRF to fuse the information from multiple sensors is seldom investigated. In [54], image and LIDAR point cloud fusion was employed for semantic segmentation. However, the fusion was done in the unary classification stage and CRF is only used as a post-process of super-pixels labeling. In [22], LIDAR point cloud was firstly clustered to generate object hypotheses. Then CRF was employed to integrate the object prior and the spatial constraints for the segmentation of the pixels. In [50], the authors proposed learning classifiers for both image and LIDAR point cloud and then using CRF to integrate the observations from the camera and LIDAR. However, in these works, the CRF model is dominated by the image, and the LIDAR points are only used as an additional observation or constraint of the registered pixels to correct the unary potentials. This paper extends their work to explicitly model the contextual interaction between the neighboring LIDAR points, and the consistency constraint between the registered pixels and LIDAR points within a novel hybrid CRF framework. Our method integrates the image and LIDAR point cloud in a totally probabilistic way and thus the information from both sensors is well exploited and deeply fused.

3. Image and LIDAR point cloud alignment

In this section, we give a brief introduction to the alignment of image and LIDAR point cloud. As presented in [16], the Velodyne HDL-64E LIDAR and a camera are mounted on the roof of a vehicle and they are synchronized by a hardware trigger. Once the rolling LIDAR is facing forward, the camera gets triggered. The camera and LIDAR are cross-calibrated so that the point cloud can be aligned with the image by projecting the LIDAR points onto the image plane [16]. Denoting a 3D point in the LIDAR coordinate by $p = [x\ y\ z\ 1]^T$, it is first transformed to the camera coordinate by

$$p_c = \mathbf{R}_{rect}\mathbf{T}_{velo}^{cam}p, \tag{1}$$

where $\mathbf{T}_{velo}^{cam}$ is the transformation matrix from the LIDAR coordinate to the camera coordinate, and $\mathbf{R}_{rect}$ is the rectifying rotation matrix.

After this step, points with negative Z-value are removed. Then the remaining points can be projected onto the image plane with the projection matrix $\mathbf{P}_{rect}$ by

$$[u'\ v'\ w]^T = \mathbf{P}_{rect}[x_c\ y_c\ z_c\ 1]^T. \tag{2}$$

Then the projected pixel coordinates of the LIDAR point p can be obtained by $[u, v] = [\frac{u'}{w}, \frac{v'}{w}]$. Note that the points that project out of the field of view (FOV) of the image are also discarded. Fig. 1 shows a point cloud captured by LIDAR and an image captured by camera in a typical road scene and the alignment of the image and point cloud. From the tree trunks in the fused view, the image and the LIDAR point cloud can be seen to be well aligned.

4. Hybrid-CRF-based camera-LIDAR fusion for road detection

In this paper, road detection is formulated as a binary labeling problem, i.e., labeling the perception data as either road (1) or background (0). The CRF-based labeling framework is adopted. The proposed method is a multi-sensor extension to the classical pairwise CRF. In this section, we first briefly introduce the CRF-based labeling framework. Then we show how to fuse the information of the image and LIDAR point cloud deeply with a novel hybrid CRF model.

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048

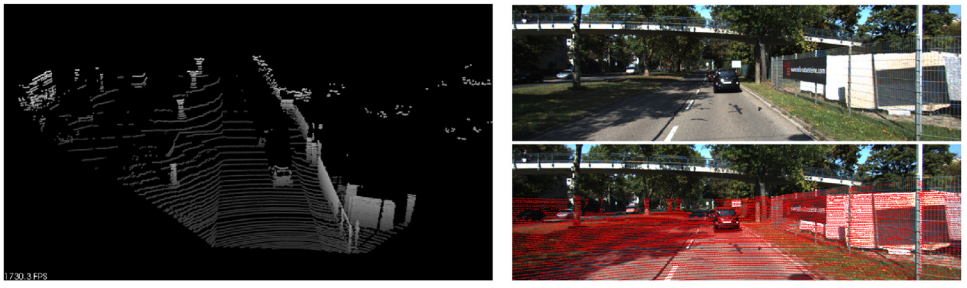

Fig. 1. On the left is a LIDAR point cloud (only the data overlapped with the FOV of the image are shown, grayscaled by height). On the top right is the corresponding image captured by camera. On the bottom right is the result of image and LIDAR point cloud fusion (note the well aligned tree trunks).

4.1. CRF-based labeling

Conditional Random Field (CRF) is a kind of probabilistic graphical model which is widely used for solving labeling problems. Formally, let $\mathbf{X} = \{X_1, X_2, \ldots X_N\}$ be the discrete random variables to be inferred from observation $\mathbf{Y}$. Each of the random variables can take a label from a predefined set $\mathcal{L} = \{l_1, l_2, \ldots l_k\}$. Any possible assignment of all the random variables is called a labeling and is denoted as $\mathbf{x}$ which can take values from $\mathbf{L} = \mathcal{L}^N$. The task is to infer the most probable labeling given the observation: $\mathbf{x}^* = \max_{\mathbf{x}\in\mathbf{L}} Pr(\mathbf{x}|\mathbf{Y})$.

A CRF is a probabilistic graphical model defined over $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{X_1, X_2, \ldots X_N\}$ and $\mathcal{E}$ defines the neighboring or connectivity between the random variables. A clique $c \in C_\mathcal{G}$ is a set of random variables $\mathbf{X}_c$ which are conditionally dependent on each other. According to the Hammersley–Clifford theorem [18], the posterior distribution $Pr(\mathbf{x}|\mathbf{Y})$ over the labelings of the CRF is a Gibbs distribution and can be written as:

$$Pr(\mathbf{x}|\mathbf{Y}) = \frac{1}{Z(\mathbf{Y})} \exp\left(-\sum_{c\in C_\mathcal{G}} \psi_c(\mathbf{x}_c|\mathbf{Y})\right), \tag{3}$$

where $\psi_c(\mathbf{x}_c|\mathbf{Y})$ is the potential function defined over the clique $\mathbf{x}_c$; $C_\mathcal{G}$ is the set of cliques, and $Z(\mathbf{Y})$ is the partition function. Therefore, maximizing the probability $Pr(\mathbf{x}|\mathbf{Y})$ equals minimizing the Gibbs energy function:

$$\min_{\mathbf{x}} E(\mathbf{x}|\mathbf{Y}) = \sum_{c\in C_\mathcal{G}} \psi_c(\mathbf{x}_c|\mathbf{Y}). \tag{4}$$

For notational convenience, the conditioning on $\mathbf{Y}$ is dropped in the rest of this paper.

In computer vision, the mostly used CRF model is the pairwise CRF which only considers the unary and pairwise cliques:

$$\min_{\mathbf{x}} E(\mathbf{x}) = \sum_{i\in\mathcal{V}} \psi_i(x_i) + \sum_{(i,j)\in\mathcal{E}} \psi_{ij}(x_i, x_j). \tag{5}$$

Using the above CRF-based labeling framework, graphical models in the image and point cloud domains can be built and the pixels or the LIDAR points can be labeled via model inference. CRF has been successfully applied in various labeling problems for its ability in modeling contextual interaction. However, each sensing modality has its inherent drawbacks. For example, image quality can be seriously affected by illumination. As shown in Fig. 9, the large shadow presented on the road makes it hard to recognize. Fusing the information from both sensors can overcome the drawbacks of a single sensor and improve the performance. To exploit the advantage of CRF and sensor fusion, in this paper, the CRF-based labeling framework is extended to integrate the image and point cloud in a hybrid CRF model.

4.2. Hybrid CRF with camera-LIDAR fusion

The details of the proposed hybrid CRF are as follows. After aligning the image and LIDAR point cloud with the method introduced in Section 3, we take the labels of the image pixels (P) and the LIDAR points (L) which project onto the field of view of the image as random variables. Because the road detection is formulated as a two-class labeling problem, each random variable can take a value from $\mathcal{L} = \{0, 1\}$. For the neighboring relationship, three types of edges are considered: (i) The first is the pixel to pixel edges (E_{PP}) which connect a pixel with its 8-neighboring pixels. (ii) The second is the edges between the neighboring LIDAR points (E_{LL}). Practically, either the K-nearest neighbor (K-NN) approach or the ϵ-neighbor approach in the 3D Euclidean space can be used. In this paper, the K-NN approach is adopted with $K = 6$. (iii) The last type (E_{PL}) is the cross-modal edges between the aligned LIDAR points and the corresponding pixels, i.e., an edge is added between each LIDAR point and the pixel which the LIDAR point is projected on. The graphical model is also illustrated in Fig. 2.

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048

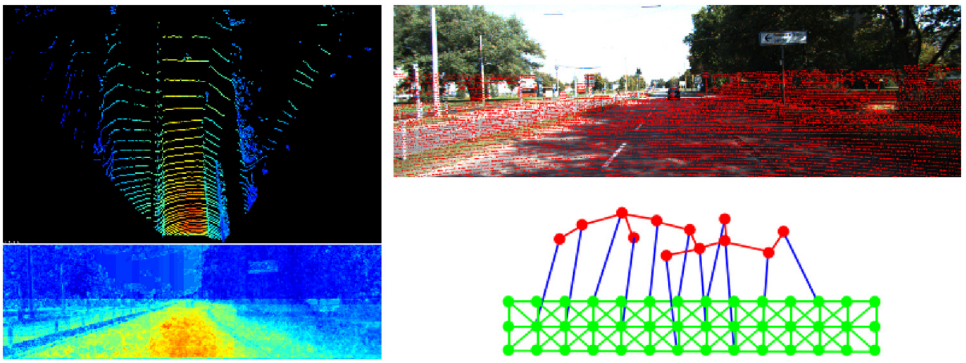

Fig. 2. Illustration of the proposed model. On the top left is the probabilistic output of the point cloud classifier. On the bottom left is the probabilistic output of the pixel classifier. On the top right is the fused view of the image and LIDAR point cloud. On the bottom right is the graph structure of the hybrid CRF: The green nodes represent the image pixels, the red nodes stand for LIDAR points, and the three kinds of edge E_{PP} (pixel to pixel), E_{LL} (LIDAR point to LIDAR point) and E_{PL} (pixel to LIDAR point) are shown in green, red and blue, respectively (See text for details). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Formally, the energy function to be minimized of the hybrid CRF is modeled as:

$$\begin{aligned}\min_{\mathbf{x}} E(\mathbf{x}) = & \overbrace{\sum_{i\in P}\psi_i^P(x_i) + \sum_{(i,j)\in E_{PP}}\psi_{ij}^P(x_i,x_j)}^{E_P(\mathbf{x}_P)} \\ & +\gamma\cdot\underbrace{\left(\sum_{i\in L}\psi_i^L(x_i) + \sum_{(i,j)\in E_{LL}}\psi_{ij}^L(x_i,x_j)\right)}_{E_L(\mathbf{x}_L)} \\ & + \sum_{(i,j)\in E_{PL}}\psi_{ij}^C(x_i,x_j).\end{aligned} \tag{6}$$

In this hybrid CRF model, there are two sub-CRFs that are built in the image and point cloud domains, respectively. The sub-energy functions E_P and E_L are the same as those of the conventional pairwise CRF models, and the strengths of the two sub-models are balanced by parameter γ. Practically, for the pixel-based sub-CRF model, the framework of Shotton's TextonBoost [44] is mostly adopted. In other words, the unary potential takes the output of a learned classifier, and the pairwise potential takes the pixel contrast sensitive Potts model. In this paper, the unary potential term $\psi_i^P(x_i)$ takes the negative log-likelihood predicted by the boosted pixel classifier:

$$\psi_i^P(x_i) = -\log p(x_i). \tag{7}$$

The pairwise potential term $\psi_{ij}^P(x_i,x_j)$ penalizes the neighboring pixels which take different labels as follows:

$$\psi_{ij}^P(x_i,x_j) = \begin{cases} 0, & \text{if } x_i = x_j \\ \lambda\cdot\frac{1}{\text{dist}(i,j)}\cdot\exp\left(-\frac{\|I_i-I_j\|^2}{2\beta}\right), & \text{otherwise.}\end{cases} \tag{8}$$

where I_i is the vector of the RGB values of the pixel i; β is expectation of $\|I_i - I_j\|^2$ over an image sample, and dist(i, j) is the Euclidean distance between the pixel site i and j. Note that the 8-neighborhood system is adopted in this paper, thus dist(i, j) equals 1 for horizontally or vertically connected neighbors and $\sqrt{2}$ for diagonally connected ones. λ is the parameter which controls the strength of the pairwise term.

For the LIDAR-point-based sub-CRF model, the unary potential of LIDAR points $\psi_i^L(x_i)$ also takes the negative log-likelihood predicted by the learned classifier for class x_i as:

$$\psi_i^L(x_i) = -\log p\prime(x_i). \tag{9}$$

For the LIDAR point to LIDAR point pairwise potential, a distance aware Potts model is adopted. The neighboring points with smaller distance are considered to be more likely to have the same label. In this paper, the potential term is formulated as:

$$\psi_{ij}^L(x_i,x_j) = \begin{cases} 0, & \text{if } x_i = x_j \\ \zeta\cdot\exp\left(-\|p_i - p_j\|^2\right), & \text{otherwise,}\end{cases} \tag{10}$$

where p_i is the 3D location vector of the LIDAR point i, and ζ is the parameter controlling the strength of enforcing the close points to take the same labels.

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048

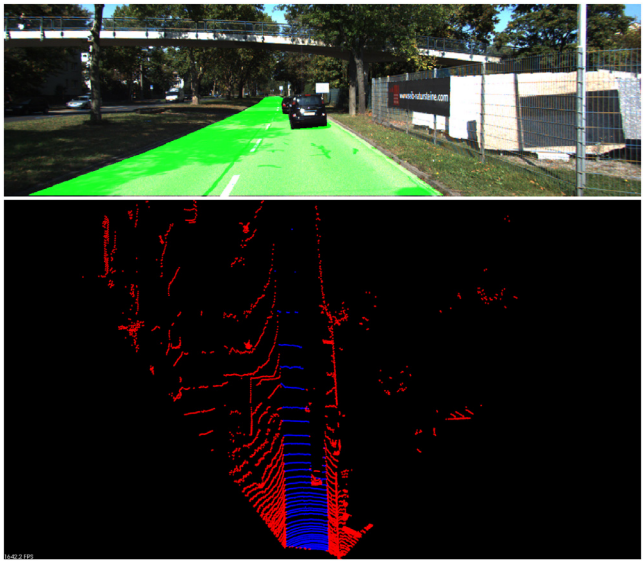

Fig. 3. Labeling of the image and point cloud. Top: the manually labeled image. Bottom: the labeling transferred from the image to point cloud.

For the pixel to LIDAR point pairwise potential, the basic Potts model is adopted:

$$\psi_{ij}^{C}(x_i, x_j) = \begin{cases} 0, & \text{if } x_i = x_j \\ \eta, & \text{otherwise,} \end{cases} \tag{11}$$

where η is the parameter controlling the strength of constraining the aligned LIDAR point and pixel to take the same labels.

Note that although the camera and LIDAR are temporally synchronized and cross calibrated, there still are minor mismatches in registration. However, the mismatches mostly exist near the edge of the objects, while in the flat road areas, the mismatches are negligible. Besides, the cross-modal potential term imposes a soft constraint on the label consistency of the aligned LIDAR points and pixels, instead of a hard one. The inference of the overall hybrid CRF model will find a solution which balances all the potential terms. Therefore, the mismatches existing in registration are acceptable in the proposed model.

4.3. Model optimization

The energy function of the proposed hybrid CRF model is sub-modular and the exact optimal labeling can be inferred efficiently by graph cuts [23]. In this paper, the fast max-flow algorithm[1] proposed by Boykov and Kolmogorov [6] is employed to solve the energy minimization problem.

5. Unary classifiers training

In the last section, the details of the proposed hybrid CRF model are introduced. In the model, the unary potentials of the pixels and LIDAR points are derived from the outputs of the learned classifiers. The performance of the unary classifiers plays an important role in the whole model. In the literature, various machine learning methods have been applied to road detection. Boosting is a famous method which is efficient in training and generalizes well to unknown samples [13]. For these desiderata, boosting is utilized for unary classifiers training.

5.1. Training samples labeling

Because images and point clouds are used at the same time, it is necessary to label both the images and the point clouds. While labeling the images can be rather easy, labeling the point clouds is much more labor-intensive. Considering that the images and the point clouds have been aligned, it is possible to label the images only and then transfer the label information to the corresponding point clouds.

For example, the KITTI-ROAD dataset [14] employed in our experiments is labeled only in the image domain. The aforementioned scheme is used to generate ground-truth labels for the LIDAR points. Fig. 3 shows an example of the labeling results.

[1] http://vision.csd.uwo.ca/code/maxflow-v3.01.zip.

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048

5.2. Feature extraction

5.2.1. Image features

The *Texture Filter Bank Response, Local Binary Pattern, Dense HOG* and *Color* are extract pixel-wise as the image features. The *Location* cues are also included in the features.

- *Texture Filter Bank Response* The images are converted to the CIE-*Lab* color space and then the filter bank is applied to the gray scale image or each channel of the CIE-*Lab* image. In practice, a Gaussian filter is applied to each channel while the horizontal and vertical Gaussian Derivative filters and the Laplacian of Gaussian filter are applied to the gray scale image. Therefore, for a given scale σ, a 6-dimensional feature vector is obtained for each pixel. In this paper, three scales are employed and thus an 18-dimensional filter bank response is obtained for each pixel.
- *Local Binary Pattern* The 8-connected neighboring local binary pattern feature is extracted to describe the local texture additionally.
- *Dense HOG* The dense Histogram of Oriented Gradients is calculated for 9 directions.
- *Color* The RGB channels of each pixel are included in the features.
- *Location* The location of the pixel is a useful cue for road detection because the road always appears at the lower part of the image. Thus, the 2D normalized x and y coordinates of the pixel are also used as part of the features.

Finally, a 40-dimensional feature vector is obtained for each pixel in the image. The feature vector is then fed into the classifier to get the probability of being either road or background for each pixel.

5.2.2. Point cloud features

For the point cloud features, several commonly used simple geometric features are used:

- *3D Position* The 3D position is represented with the 3D coordinates normalized with the distance.
- *Spectral Features* Denoting $\lambda_0 < \lambda_1 < \lambda_2$ as the eigenvalues of the scatter matrix M estimated from the local neighborhood of point p, $\{\sigma_p = \lambda_0, \sigma_s = \lambda_1 - \lambda_0, \sigma_l = \lambda_2 - \lambda_1\}$ are extracted as the spectral features [40].
- *Directional Features* Local tangent $\vec{v_t}$ and normal $\vec{n_t}$ vectors are estimated by the principal and least eigenvectors of M and these vectors are used as the directional features.

Thus, a 12-dimensional feature vector is obtained for each LIDAR point.

5.3. Classifier training

With data labeled and features extracted, classifiers can be trained. In this paper, the boosted decision tree is chosen as the classifier for both the images and the point clouds. Boosting iteratively learns a strong classifier as a sum of weak classifiers. In this work, the decision tree classifier with depth d is taken as the weak classifier and AdaBoost is taken as the boosting algorithm. Denoting the feature vector as $\mathbf{v}$, each weak classifier $h_i(\mathbf{v})$ maps the feature to a binary prediction. The learned strong classifier $H(\mathbf{v})$ after N iterations of AdaBoost is a weighted sum of the weak classifiers:

$$H(\mathbf{v}) = \sum_{i}^{N} \alpha_i \cdot h_i(\mathbf{v}). \tag{12}$$

The strong classifier outputs a confidence value for each testing feature to take label 1 (road) and label 0 (background):

$$c(x|\mathbf{v}) = \sum_{i}^{N} \alpha_i \cdot h_i(\mathbf{v}), \qquad x \in \{0, 1\}. \tag{13}$$

Then the confidence values can be reinterpreted as probabilities by:

$$p(x = 1|\mathbf{v}) = \frac{c(x = 1|\mathbf{v})}{c(x = 1|\mathbf{v}) + c(x = 0|\mathbf{v})} \tag{14}$$

and

$$p(x = 0|\mathbf{v}) = \frac{c(x = 0|\mathbf{v})}{c(x = 1|\mathbf{v}) + c(x = 0|\mathbf{v})}. \tag{15}$$

These probabilities can be used to obtain the unary potentials in the hybrid CRF model.

6. Experiments

6.1. Dataset

In this section, we conduct some experiments on the publicly available KITTI-ROAD dataset [14] to validate the performance of the proposed approach. The KITTI-ROAD dataset contains sensor data captured with car-mounted hardware-synchronized cameras and Velodyne HDL-64E LIDAR. The cross calibration parameters are also offered to get the LIDAR

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048

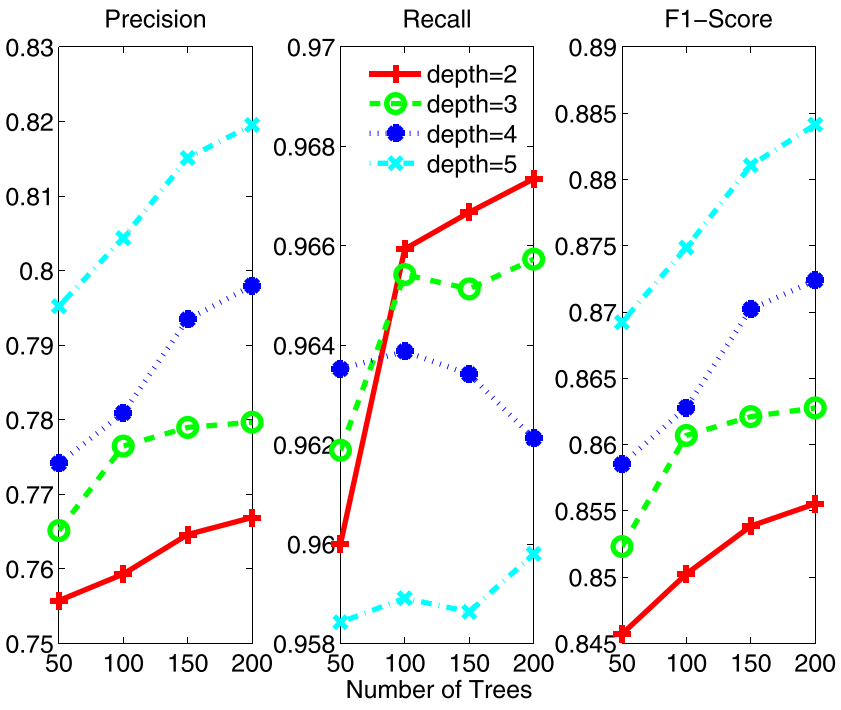

Fig. 4. Classification performance with different numbers of trees and depths.

point clouds registered with the images. The whole dataset contains about 600 frames of recording captured from five different days with relatively low traffic density. The data is organized into three sub-categories according to the driving environments: urban marked (UM), urban marked multi-lane (UMM) and urban unmarked (UU). Each of them consists of about 100 training frames and 100 testing frames. The annotated images are offered for the training frames, while the ground-truth for the testing data is not publicly available and one needs to upload the results to get evaluated online. The annotations contain the road area and the ego-lane. In this paper, only the road detection is studied, while the ego-lane information is ignored. Two kinds of metrics are offered for evaluation: one is the pixel-based evaluation in the perspective view and the other is the behavior-based evaluation in the bird's eye view (BEV). The performance indices include false positive rate (FPR), false negative rate (FNR), precision (PRE), recall (REC), and F1-score. Considering some methods output confidence maps, the maximum F1-score (MaxF) and the average precision (AP) [14] are also computed in the official development kit [1]. Because our method output binary prediction, MaxF is equal to F1-score and AP is not quite suitable for evaluating our method. Therefore, in this paper, for the results evaluated in the official way, MaxF and AP are listed for the sake of consistency with the leaderboard [1]. Otherwise, AP is omitted and F1-score is used instead of MaxF.

6.2. Parameter settings

6.2.1. Classifiers

We first test the performance of the boosted decision tree classifiers. As is known, the weak classifier and the running rounds of AdaBoost are the two main impact factors of boosting. Taking the pixel-wise classification of the UM subset as an example, we conduct 2-fold cross validation with different configurations of the decision tree depths and rounds of AdaBoost (i.e. the numbers of trees). Fig. 4 shows the changes of Precision, Recall and F1-score under different parameter configurations. From the figure, we can observe that better F1-scores can be achieved when we employ deeper weak classifiers and run for more rounds. However, the corresponding running time increases too. Therefore, it is necessary to find a balance between performance and efficiency. The testing time with respect to different numbers of trees and depths is shown in Fig. 5. Considering both the performance and the efficiency, we take 100 depth-4 decision trees as the strong classifiers for the pixels and LIDAR points.

6.2.2. Hybrid CRF parameters

In the hybrid CRF model, there are several parameters controlling the importance of the different kinds of potential terms. These parameters have a major impact on the performance of the proposed method. Again, we employ 2-fold cross validation to search for the best parameters. The hybrid CRF model is built based on two unimodal pairwise CRF models. The parameters are tuned as follows: Firstly, we find the best pixel to pixel pairwise weight λ in the pixel-based pairwise CRF model. Then, the parameter of LIDAR point to LIDAR point pairwise term ζ is tuned in the LIDAR-point-based pairwise

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048

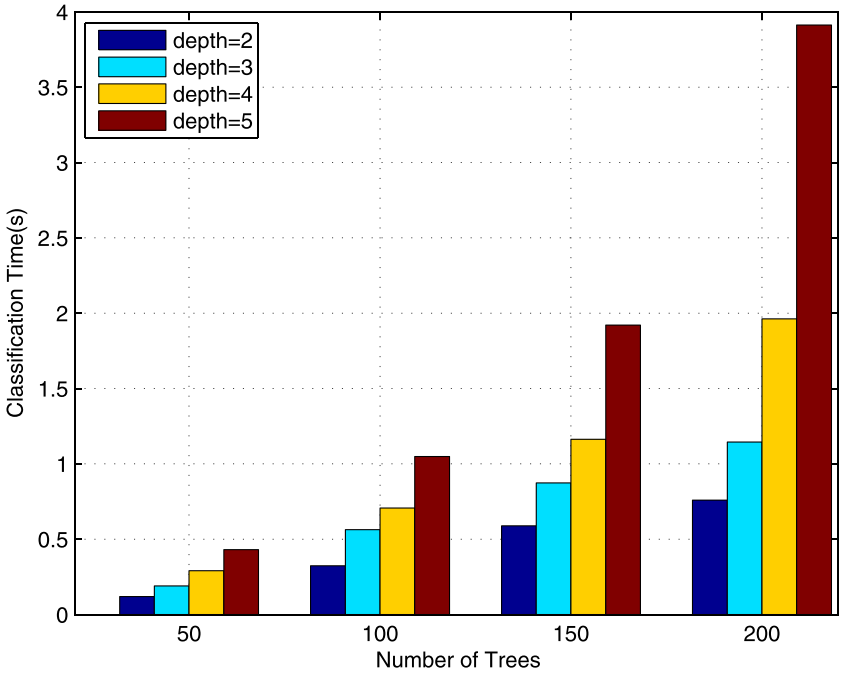

Fig. 5. Running time of pixel classification with different numbers of trees and depths.

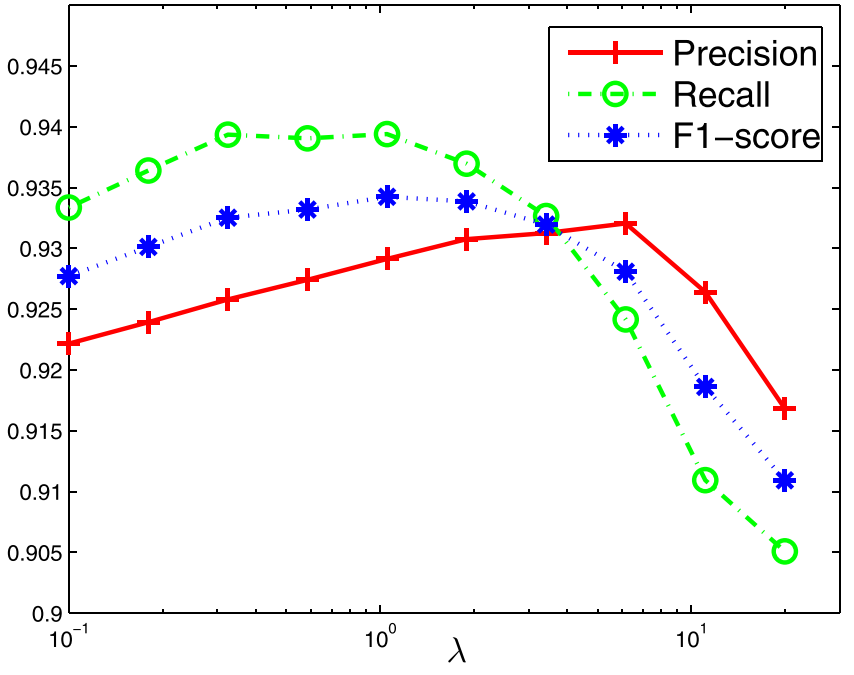

Fig. 6. Results of cross validation under different λ.

CRF model. Then the best parameters λ and ζ are fixed, and the remaining two parameters γ and η are tuned in the hybrid CRF model.

Taking the UM subset as an example, we first tune parameters λ and ζ separately within the pixel-based pairwise CRF and LIDAR-point-based pairwise CRF. We use Precision, Recall and F1-score to evaluate the performance of different parameter settings. The results are shown in Figs. 6 and 7. With the results shown in the figures, we can choose the parameters λ and ζ with the best F1-score.

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048

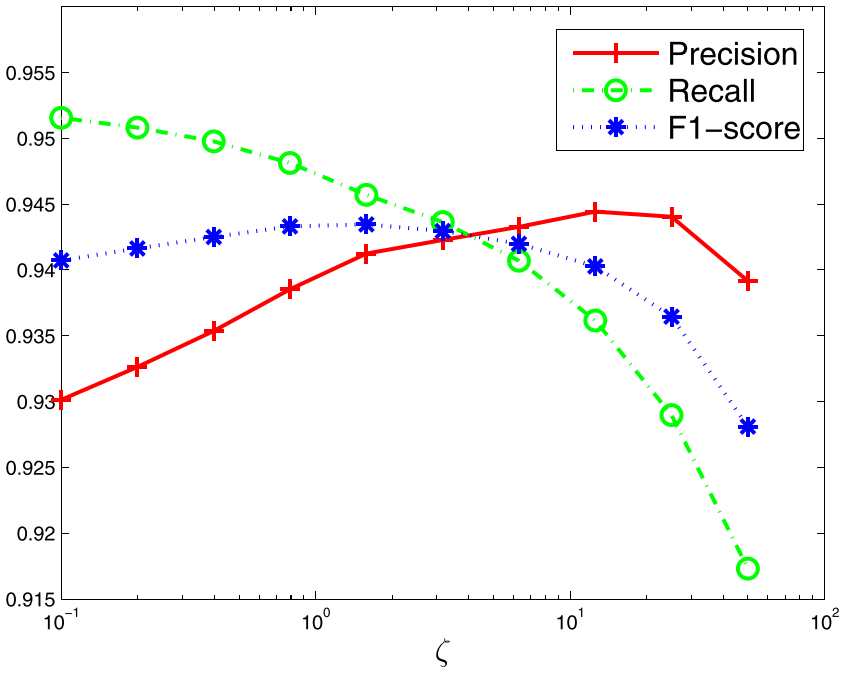

Fig. 7. Results of cross validation under different ζ.

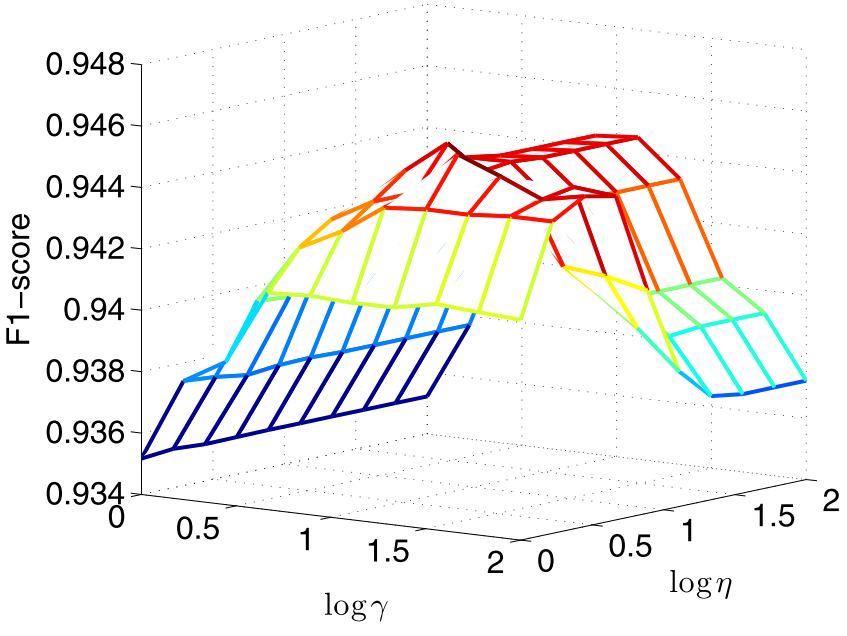

Fig. 8. F1-scores of cross validation under different γ and η.

Then these two parameters are fixed and we perform 2-fold cross validation for different settings of γ and η in the hybrid CRF model. Fig. 8 shows the F1-scores under different parameter settings. It can be seen from the figure that when the parameters grow from small values, the F1-score increases, but when the parameters grow too big, the F1-score decreases quickly. This figure can help us select the best γ and η with the highest F1-score.

6.3. Performance evaluation

To evaluate the performance of the proposed model, we conduct several comparative experiments on the KITTI-ROAD dataset. In the first stage, we randomly divide the training images with ground-truth provided into two equally numbered

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048

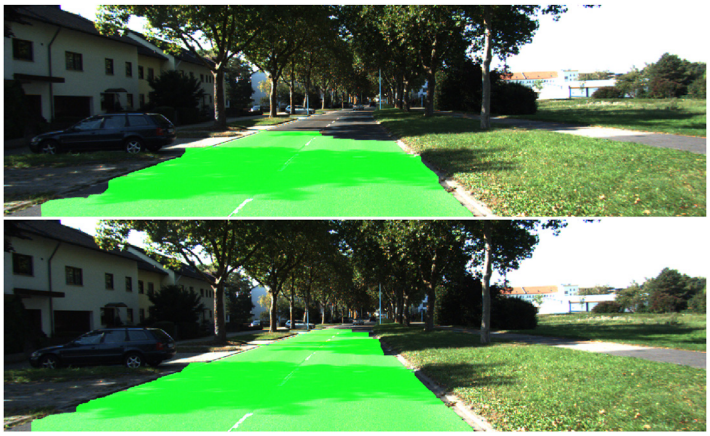

Fig. 9. Comparison of road detection by pixel-based CRF (top) and the proposed HybridCRF (bottom). The green areas denote the detected roads. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Comparison on UM (perspective view).

Algorithm	MaxF	AP	PRE	REC	FPR	FNR
Pixel classifier	89.93	77.93	91.54	88.38	1.64	11.62
Pixel-wise CRF	92.52	86.15	93.10	91.95	1.37	8.05
HybridCRF	**94.47**	**87.38**	**94.44**	**95.84**	**1.12**	**5.50**

Table 2
Comparison on UMM (perspective view).

Algorithm	MaxF	AP	PRE	REC	FPR	FNR
Pixel classifier	92.45	85.75	91.94	92.97	2.55	7.03
Pixel-wise CRF	94.49	87.28	93.62	95.37	2.03	4.63
HybridCRF	**95.53**	**88.50**	**94.97**	**96.10**	**1.59**	**3.90**

Table 3
Comparison on UU (perspective view).

Algorithm	MaxF	AP	PRE	REC	FPR	FNR
Pixel classifier	86.15	73.66	86.86	85.45	2.15	14.55
Pixel-wise CRF	90.96	79.23	**93.66**	88.42	**1.00**	11.58
HybridCRF	**92.64**	**85.96**	93.13	**92.16**	1.13	**7.84**

sets: one used for training and the other used for testing. As the proposed method is a composite model of two sub-CRF models, we firstly compare our method with the unimodal sub-CRF models, namely, the pixel-based CRF and the LIDAR-point-based CRF. Because the ground-truth of the KITTI-Road dataset is provided only in the image domain, it is unsuitable to evaluate the LIDAR-based methods. Thus, we compare the pixel-based CRF and the LIDAR-point-based CRF with our method separately.

For the pixel-based CRF, we can evaluate it and compare it with our method in the image domain using the official development kit. Fig. 9 shows an example of the results obtained by the pixel-wise CRF and the proposed hybrid CRF, the overlapped green area denotes the detected road. From the figure, it can be seen that the result of the pixel-based pairwise CRF is affected by the shadow of the trees projected on the road, while in the proposed hybrid CRF, with LIDAR point cloud fused, the impact of shadow has been reduced.

Then quantitative evaluation is performed with the indices introduced in [14]. Additionally, we take the output of the pixel classifier (it is also a special case of pixel-based CRF with $\lambda = 0$) as the baseline. Note that the three subsets are treated separately. In other words, no data from one subset are used for training or testing of the other subsets. The evaluation is performed in the perspective view. The results on the UM, UMM and UU subsets are shown in Tables 1–3.

Similarly, the same comparative experiments are conducted with the LIDAR-point-based CRF. The performance is evaluated in terms of LIDAR point-wise accuracy and the ground-truth labels are obtained by transferring the labels from the ground-truth images to the LIDAR points by registration. The outputs of the boosted decision tree classifier are also taken as another baseline. The results tested on the UM, UMM and UU subsets are listed in Tables 4–6.

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048

Table 4
Comparison on UM (point cloud).

Algorithm	F1-score	PRE	REC	FPR	FNR
Point classifier	94.74	93.53	95.98	3.36	4.02
Point-wise CRF	95.28	94.29	96.29	2.95	3.71
HybridCRF	**96.06**	**95.61**	**96.51**	**2.24**	**3.49**

Table 5
Comparison on UMM (point cloud).

Algorithm	F1-score	PRE	REC	FPR	FNR
Point classifier	92.68	94.62	90.82	8.25	5.38
Point-wise CRF	93.64	96.00	91.39	7.80	4.00
HybridCRF	**94.86**	**96.50**	**93.27**	**6.01**	**3.50**

Table 6
Comparison on UU (point cloud).

Algorithm	F1-score	PRE	REC	FPR	FNR
Point classifier	91.82	90.33	93.37	3.79	6.63
Point-wise CRF	93.27	92.45	**94.11**	2.92	**5.89**
HybridCRF	**94.39**	**94.75**	94.02	**1.98**	5.98

Table 7
Results of online evaluation on KITTI-UM (BEV).

Algorithm	MaxF	*AP*	PRE	REC	FPR	FNR
HIM [39]	90.07	79.98	90.79	89.35	4.13	10.65
SPRAY [24]	88.14	**91.24**	88.60	87.68	5.14	12.32
BM [49]	78.90	66.06	69.53	91.19	18.21	8.81
ProbBoost [48]	87.48	80.13	85.02	90.09	7.23	9.91
HistonBoost [47]	83.68	72.79	82.01	85.42	8.54	14.58
RES3D-Velo [43]	83.81	73.95	78.56	89.80	11.16	10.20
FusedCRF [50]	89.55	80.00	84.87	**94.78**	7.70	**5.22**
PGM-ARS [41]	80.97	69.11	77.51	84.76	11.21	15.24
CB [36]	88.89	82.17	87.26	90.58	6.03	9.42
StixelNet [27]	85.33	72.14	81.21	89.89	9.48	10.11
NNP [9]	90.50	87.95	**91.43**	89.59	**3.83**	10.41
SRF [51]	76.43	83.24	75.53	77.35	11.42	22.65
FCN-LC [35]	89.36	78.80	89.35	89.37	4.85	10.63
MAP [25]	87.33	89.62	85.77	88.95	6.73	10.63
HybridCRF(Ours)	**90.99**	85.26	90.65	91.33	4.29	8.67

From the results shown in the tables, it can be seen that with contextual information modeled, the pairwise CRF model improves the performance over the pixel or LIDAR-point-based classifiers. Apart from the contextual information, the proposed hybrid CRF also fuses the multi-sensor information in an integrated probabilistic model. Therefore, the inferred labels are more accurate than the unimodal CRF models.

With the superiority over the unimodal pairwise CRF models validated, we conduct some comparative experiments with the recently developed ones. In these experiments, all the training data with the ground-truth provided in each subset is used to learn the classifiers for that subset. The parameters take the optima obtained by cross validation in the previous subsection. The results are transformed to the bird's eye view (BEV) and then submitted to the website for evaluation. Fig. 10 shows some examples of testing images evaluated in the BEV. More results can be found on the website [1].

We compare the proposed approach with the high ranking methods on the leaderboard [1], including HIM [39], SPRAY [24], BM [49], ProbBoost [48], HistonBoost [47], RES3D-Velo [43], FusedCRF [50], PGM-ARS [41], CB [36], StixelNet [27], NNP [9], SRF [51], FCN-LC [35] and MAP [25]. Among these methods, there are camera-LIDAR fusion methods (RES3D-Velo and FusedCRF) and stereo-vision-based methods (HistonBoost, ProbBoost and NNP). PGM-ARS and FusedCRF are methods which also take advantage of CRF. FCN-LC and MAP are deep-learning-based methods which exploit the powerful fully convolutional networks [34].

The results on the UM, UMM, UU subsets and the average results are listed in Tables 7–10. The best performance indices are shown in bold. Note that the average precisions (AP) of our method are far worse than the best. The reason is that the AP is designed for evaluating probabilistic predictions, while our method outputs binary predictions. Therefore, it is not suitable to be evaluated in that metric. This point is also stated in [49]. For more detailed information about the results of these methods, we refer the readers to the website [1]. From the results, it can be seen that the proposed method is competitive. In particular, it achieves the best F1-scores on the UM and UU subsets. However, the results on the UMM

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048

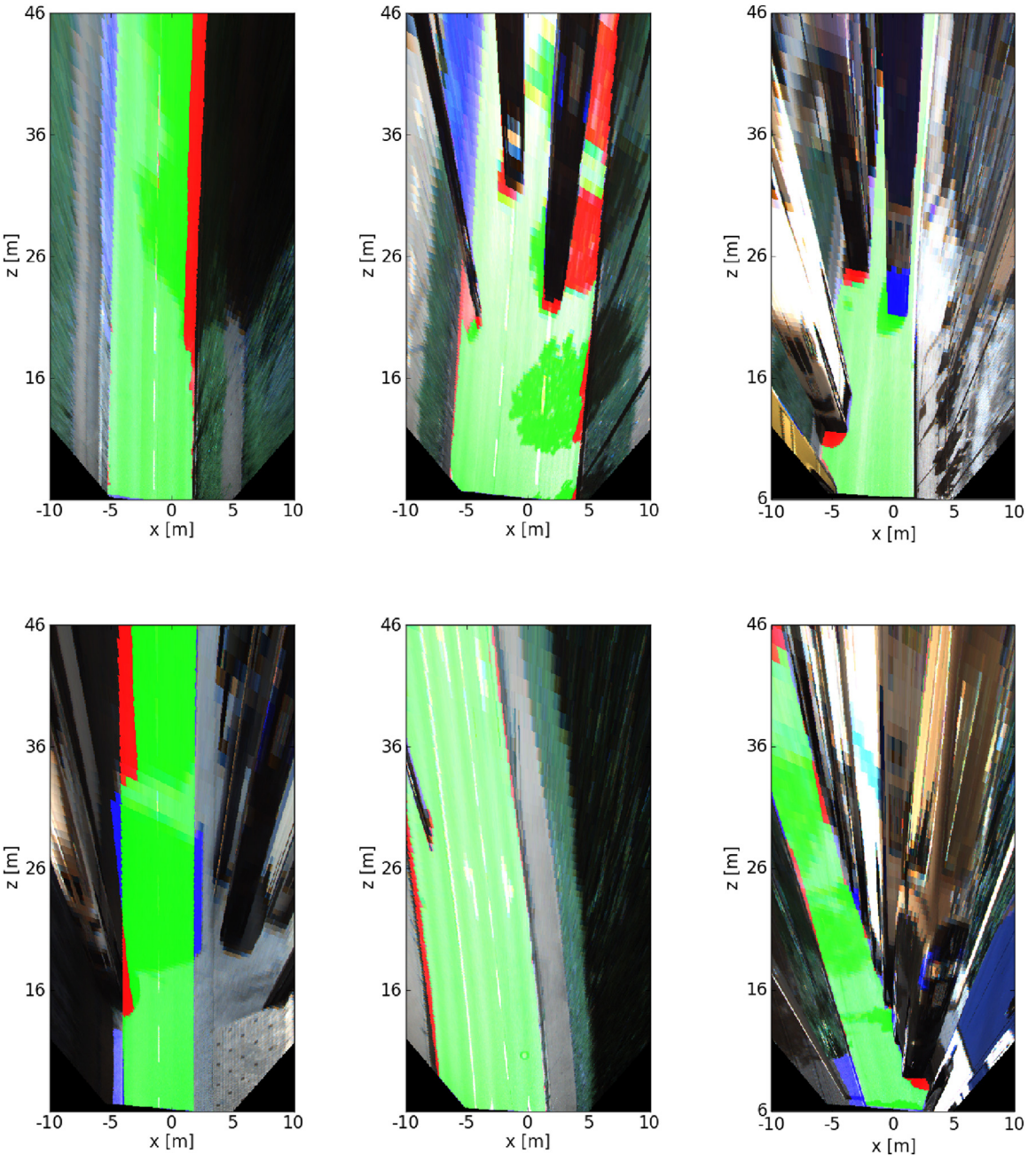

Fig. 10. Road detected in the bird's eye view. Here, the red denotes false negatives; the blue areas correspond to false positives, and the green represents true positives. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

subset is not as good as. That may be because the LIDAR points become sparser in the UMM scenes which have much wider roads. Therefore, the LIDAR information becomes less discriminative, especially between the roads and sidewalks. Nevertheless, the proposed method achieves the best F1-score in these algorithms on the whole dataset.

In addition, the computational time is an important factor in developing and selecting road detection methods. For the proposed method, the computation cost consists of feature extraction and classification of the pixels and LIDAR points, graph construction and cutting. The proposed algorithm is implemented in standard, single-threaded C++ and tested on a standard PC with 8GB memory and an Intel(R) Core(TM) i5-3230M CPU clocked at 2.6 GHz. The average computational time tested on the KITTI-ROAD dataset is approximately 1.5 s. Although the current implemented version has not been used in real time, we can use parallel computing and sub-sample the image for accelerating when in real-time application.

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048

Table 8
Results of online evaluation on KITTI-UMM (BEV).

Algorithm	MaxF	*AP*	PRE	REC	FPR	FNR
HIM [39]	93.55	90.38	**94.18**	92.92	6.31	7.08
SPRAY [24]	89.69	**93.84**	89.13	90.25	12.10	9.75
BM [49]	89.41	80.61	83.43	**96.30**	21.02	**3.70**
ProbBoost [48]	91.36	84.92	88.18	94.78	13.97	5.22
HistonBoost [47]	88.73	81.57	84.49	93.42	18.85	6.58
RES3D-Velo [43]	90.60	85.38	85.96	95.78	17.20	4.22
FusedCRF [50]	89.51	83.53	86.64	92.58	15.69	7.42
PGM-ARS [41]	91.76	84.80	88.05	95.80	14.30	4.20
CB [36]	90.55	85.40	92.75	88.45	7.60	11.55
StixelNet [27]	93.26	87.15	90.63	96.06	10.92	3.94
NNP [9]	91.34	88.65	91.07	91.60	9.87	8.40
SRF [51]	90.77	92.44	89.35	92.23	12.08	7.77
FCN-LC [35]	**94.09**	90.26	94.05	94.13	6.55	5.87
MAP [25]	89.97	92.14	87.47	92.62	14.58	7.38
HybridCRF(Ours)	91.95	86.44	94.01	89.98	**6.30**	10.02

Table 9
Results of online evaluation on KITTI-UU (BEV).

Algorithm	MaxF	*AP*	PRE	REC	FPR	FNR
HIM [39]	85.76	76.18	**87.65**	83.95	**3.86**	16.05
SPRAY [24]	82.71	**87.19**	82.16	83.26	5.89	16.74
BM [49]	78.43	62.46	70.87	87.80	11.76	12.20
ProbBoost [48]	80.76	68.70	85.25	76.72	4.33	23.28
HistonBoost [47]	74.19	63.01	77.43	71.22	6.77	28.78
RES3D-Velo [43]	83.63	72.58	77.38	90.97	8.67	9.03
FusedCRF [50]	84.49	72.35	77.13	**93.40**	9.02	**6.60**
PGM-ARS [41]	79.94	67.77	77.37	82.67	7.88	17.33
CB [36]	86.13	75.21	86.47	85.80	4.38	14.20
StixelNet [27]	86.06	72.05	82.61	89.82	6.16	10.18
NNP [9]	85.55	76.90	85.36	85.75	4.79	14.25
SRF [51]	76.07	79.97	71.47	81.31	10.57	18.69
FCN-LC [35]	86.27	75.37	86.65	85.89	4.31	14.11
MAP [25]	84.44	87.17	83.66	85.23	5.42	14.77
HybridCRF(Ours)	**88.53**	80.79	86.41	90.76	4.65	9.24

Table 10
Average results of online evaluation on KITTI-ROAD (BEV).

Algorithm	MaxF	*AP*	PRE	REC	FPR	FNR
HIM [39]	90.64	81.42	**91.62**	89.68	**4.52**	10.32
SPRAY [24]	87.09	**91.12**	87.10	87.08	7.10	12.92
BM [49]	83.47	72.23	75.90	92.72	16.22	7.28
ProbBoost [48]	87.78	77.30	86.59	89.01	7.60	10.99
HistonBoost [47]	83.92	73.75	82.24	85.66	10.19	14.34
RES3D-Velo [43]	86.58	78.34	82.63	90.92	10.53	9.08
FusedCRF [50]	88.25	79.24	83.62	**93.44**	10.08	**6.56**
PGM-ARS [41]	85.69	73.83	82.34	89.33	10.56	10.67
CB [36]	88.97	79.69	89.50	88.44	5.71	11.56
StixelNet [27]	89.12	81.23	85.80	92.71	8.45	7.29
NNP [9]	89.68	86.50	89.67	89.68	5.69	10.32
SRF [51]	82.44	87.37	80.60	84.36	11.18	15.64
FCN-LC [35]	90.79	85.83	90.87	90.72	5.02	9.28
MAP [25]	87.80	89.96	86.01	89.66	8.04	10.34
HybridCRF(Ours)	**90.81**	86.01	91.05	90.57	4.90	9.43

7. Conclusions and future work

This paper proposed a new road detection method based on sensor fusion of a monocular camera and a multi-layer LIDAR. The information from the two sensors is jointly modeled in a hybrid conditional random field in which the labels of the pixels and LIDAR points are considered as random variables and the edges consist of the connections: (i) between the neighboring pixels in the image plane, (ii) between the neighboring LIDAR points in the 3D space, and (iii) between the aligned LIDAR points and their corresponding pixels. The unary potentials of the pixels and LIDAR points are all obtained by offline learned boosted decision tree classifiers. The pairwise potentials ensure the contextual consistency in images and

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048

point clouds, as well as the cross-modal consistency between the aligned pixels and LIDAR points. The proposed method deeply fuses the information from camera and LIDAR and effectively reduces the ambiguities in road detection. Experiments tested on the KITTI-ROAD benchmark dataset show that the proposed method outperforms other recently developed ones.

In the future, we are considering employing the more powerful deep learning methods to obtain the unary potential for the hybrid CRF model to boost the performance further. We can also transplant the algorithm to parallel computing units like GPU to accelerate it. Besides, the hybrid CRF framework can be readily extended to multi-class semantic labeling. We believe this novel sensor fusion model can achieve much better performance than the image-based semantic labeling approaches.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 61375050 and 91220301, and the Marsden Fund of New Zealand (2014–2017). We would like to thank anonymous reviewers for their valuable suggestions. We would also like to thank Diana Hibbert for proofreading this manuscript.

References

[1] The KITTI vision benchmark suite, (http://www.cvlibs.net/datasets/kitti/eval_road.php). Accessed: 2017-03-01.
[2] Y. Alon, A. Ferencz, A. Shashua, Off-road path following using region classification and geometric projection constraints, in: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1, in: CVPR '06, IEEE Computer Society, Washington, DC, USA, 2006, pp. 689–696.
[3] J.M. Alvarez, Y. LeCun, T. Gevers, A.M. Lopez, Semantic road segmentation via multi-scale ensembles of learned features, in: Computer Vision–ECCV 2012. Workshops and Demonstrations, Springer Berlin Heidelberg, 2012, pp. 586–595.
[4] J.M. Alvarez, A.M. Lopez, Road detection based on illuminant invariance, IEEE Trans. Intell. Transp. Syst. 12 (1) (2011) 184–193.
[5] A. Asvadi, C. Premebida, P. Peixoto, U. Nunes, 3d lidar-based static and moving obstacle detection in driving environments: an approach based on voxels and multi-region ground planes, Rob. Auton. Syst. 83 (2016) 299–311. http://dx.doi.org/10.1016/j.robot.2016.06.007.
[6] Y. Boykov, V. Kolmogorov, An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision, IEEE Trans. Pattern Anal. Mach. Intell. 26 (2004) 1124–1137.
[7] J. Byun, K.-i. Na, B.-s. Seo, M. Roh, Drivable road detection with 3d point clouds based on the MRF for intelligent vehicle, in: L. Mejias, P. Corke, J. Roberts (Eds.), Field and Service Robotics, Springer Tracts in Advanced Robotics, vol. 105, Springer International Publishing, 2015, pp. 49–60.
[8] T. Chen, B. Dai, R. Wang, D. Liu, Gaussian-process-based real-time ground segmentation for autonomous land vehicles, J. Intell. Robot. Syst. 76 (2013) 563–582.
[9] X. Chen, K. Kundu, Y. Zhu, A. Berneshawi, H. Ma, S. Fidler, R. Urtasun, 3d object proposals for accurate object class detection, NIPS, 2015.
[10] H. Dahlkamp, A. Kaehler, D. Stavens, S. Thrun, G.R. Bradski, Self-supervised monocular road detection in desert terrain, in: Robotics Science and System Conference (RSS), 2006.
[11] E.D. Dickmanns, B.D. Mysliwetz, Recursive 3-d road and relative ego-state recognition, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 199–213.
[12] B. Douillard, J. Underwood, N. Kuntz, V. Vlaskine, A. Quadros, P. Morton, A. Frenkel, On the segmentation of 3d lidar point clouds, in: Robotics and Automation (ICRA), 2011 IEEE International Conference on, 2011, pp. 2798–2805, doi:10.1109/ICRA.2011.5979818.
[13] Y. Freund, R.E. Schapire, A short introduction to boosting, in: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, Morgan Kaufmann, 1999, pp. 1401–1406.
[14] J. Fritsch, T. Kuehnl, A. Geiger, A new performance measure and evaluation benchmark for road detection algorithms, in: International Conference on Intelligent Transportation Systems (ITSC), 2013.
[15] J. Fritsch, T. Kuehnl, F. Kummert, Monocular road terrain detection by combining visual and spatial information, IEEE Trans. Intell. Transp. Syst. 15 (4) (2014) 1586–1596, doi:10.1109/TITS.2014.2303899.
[16] A. Geiger, P. Lenz, C. Stiller, R. Urtasun, Vision meets robotics: the KITTI dataset, Int. J. Rob. Res. 32 (2013) 1231–1237.
[17] C. Guo, W. Sato, L. Han, S. Mita, D. McAllester, Graph-based 2d road representation of 3d point clouds for intelligent vehicles, in: Intelligent Vehicles Symposium (IV), 2011 IEEE, 2011, pp. 715–721, doi:10.1109/IVS.2011.5940502.
[18] J.M. Hammersley, P. Clifford, Markov fields on finite graphs and lattices(1971).
[19] Z. He, T. Wu, Z. Xiao, H. He, Robust road detection from a single image using road shape prior, in: Image Processing (ICIP), 2013 20th IEEE International Conference on, 2013, pp. 2757–2761, doi:10.1109/ICIP.2013.6738568.
[20] A.B. Hillel, R. Lerner, D. Levi, G. Raz, Recent progress in road and lane detection: a survey, Mach. Vis. Appl. 25 (3) (2014) 727–745.
[21] X. Hu, S.A.R. F., A. Gepperth, A multi-modal system for road detection and segmentation, in: Intelligent Vehicles Symposium Proceedings, 2014 IEEE, 2014, pp. 1365–1370, doi:10.1109/IVS.2014.6856616.
[22] W. Huang, X. Gong, Z. Xiang, Road scene segmentation via fusing camera and lidar data, in: 2014 IEEE International Conference on Robotics and Automation (ICRA), 2014, pp. 1008–1013, doi:10.1109/ICRA.2014.6906977.
[23] V. Kolmogorov, R. Zabih, What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. 26 (2) (2004) 147–159, doi:10.1109/TPAMI.2004.1262177.
[24] T. Kuehnl, F. Kummert, J. Fritsch, Spatial ray features for real-time ego-lane extraction, in: Proc. IEEE Intelligent Transportation Systems, 2012.
[25] A. Laddha, M.K. Kocamaz, L.E. Navarro-Serment, M. Hebert, Map-supervised road detection, in: 2016 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2016, pp. 118–123, doi:10.1109/IVS.2016.7535374.
[26] J.-F. Lalonde, N. Vandapel, D.F. Huber, M. Hebert, Natural terrain classification using three-dimensional ladar data for ground robot mobility., J. Field Rob. 23 (10) (2006) 839–861.
[27] D. Levi, N. Garnett, E. Fetaya, Stixelnet: a deep convolutional network for obstacle detection and road segmentation., in: 26th British Machine Vision Conference (BMVC), 2015.
[28] H. Liu, D. Guo, F. Sun, Object recognition using tactile measurements: kernel sparse coding methods, IEEE Trans. Instrum. Meas. 65 (3) (2016) 656–665, doi:10.1109/TIM.2016.2514779.
[29] H. Liu, Y. Liu, F. Sun, Robust exemplar extraction using structured sparse coding, IEEE Trans. Neural Netw. Learn. Syst. 26 (2015) 1816–1821.
[30] H. Liu, J. Qin, F. Sun, D. Guo, Extreme kernel sparse learning for tactile object recognition, IEEE Trans. Cybern. PP (99) (2016) 1–12, doi:10.1109/TCYB.2016.2614809.
[31] H. Liu, F. Sun, Fusion tracking in color and infrared images using joint sparse representation, Sci. China Inf. Sci. 55 (3) (2012) 590–599.
[32] H. Liu, F. Sun, D. Guo, B. Fang, Structured output-associated dictionary learning for haptic understanding, IEEE Trans. Syst. Man Cybern. (2017).
[33] H. Liu, Y. Yu, F. Sun, J. Gu, Visual-tactile fusion for object recognition, IEEE Trans. Autom. Sci. Eng. PP (99) (2016) 1–13, doi:10.1109/TASE.2016.2549552.
[34] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048

[35] C. Mendes, V. Frémont, D. Wolf, Exploiting fully convolutional neural networks for fast road detection, in: IEEE Conference on Robotics and Automation (ICRA), 2016.
[36] C.C.T. Mendes, V. Frmont, D.F. Wolf, Vision-based road detection using contextual blocks, (2015). arXiv:1509.01122.
[37] R. Mohan, Deep deconvolutional networks for scene parsing, arXiv:1411.4101(2014).
[38] F. Moosmann, O. Pink, C. Stiller, Segmentation of 3d lidar data in non-flat urban environments using a local convexity criterion, Intelligent Vehicles Symposium (IV), 2009 IEEE, 2009.
[39] D. Munoz, J.A. Bagnell, M. Hebert, Stacked hierarchical labeling, in: European Conference on Computer Vision (ECCV), 2010.
[40] D. Munoz, N. Vandapel, M. Hebert, Directional associative markov network for 3-d point cloud classification, International Symposium on 3-D Data Processing, Visualization, and Transmission (3DPVT), 2008.
[41] M. Passani, J.J. Yebes, L.M. Bergasa, Fast pixelwise road inference based on uniformly reweighted belief propagation, in: Proc. IEEE Intelligent Vehicles Symposium, 2015.
[42] P.Y. Shinzato, V.G. Jr, F.S. Osorio, D.F. Wolf, Fast visual road recognition and horizon detection using multiple artificial neural networks, in: Intelligent Vehicles Symposium Proceedings, 2012 IEEE, 2012, pp. 1090–1095, doi:10.1109/IVS.2014.6856616.
[43] P.Y. Shinzato, D.F. Wolf, C. Stiller, Road terrain detection: Avoiding common obstacle detection assumptions using sensor fusion, in: Intelligent Vehicles Symposium Proceedings, 2014 IEEE, 2014, pp. 687–692, doi:10.1109/IVS.2014.6856454.
[44] J. Shotton, J. Winn, C. Rother, A. Criminisi, Textonboost: joint appearance, shape and context modeling for multi-class object recognition and segmentation, in: Computer Vision–ECCV 2006, Springer Berlin Heidelberg, 2006, pp. 1–15.
[45] W. Tao, Y. Zhou, L. Liu, K. Li, K. Sun, Z. Zhang, Spatial adjacent bag of features with multiple superpixels for object segmentation and classification, Inf. Sci. 281 (2014) 373–385. http://dx.doi.org/10.1016/j.ins.2014.05.032.
[46] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, et al., Stanley: the robot that won the darpa grand challenge, in: The 2005 DARPA Grand Challenge, Springer, 2007, pp. 1–43.
[47] G. Vitor, A. Victorino, J. Ferreira, Comprehensive performance analysis of road detection algorithms using the common urban kitti-road benchmark, in: Workshop on Benchmarking Road Terrain and Lane Detection Algorithms for In-Vehicle Application on IEEE Intelligent Vehicles Symposium (IV), 2014b, pp. 19–24, doi:10.1109/IVS.2014.6856616.
[48] G.B. Vitor, A.C. Victorino, J.V. Ferreira, A probabilistic distribution approach for the classification of urban roads in complex environments, in: Workshop on Modelling, Estimation, Perception and Control of All Terrain Mobile Robots on IEEE International Conference on Robotics and Automation (ICRA) 2014, 2014a.
[49] B. Wang, V. Fremont, S. Rodriguez, Color-based road detection and its evaluation on the kitti road benchmark, in: Intelligent Vehicles Symposium Proceedings, 2014 IEEE, 2014, pp. 31–36, doi:10.1109/IVS.2014.6856619.
[50] L. Xiao, B. Dai, D. Liu, T. Hu, T. Wu, CRF based road detection with multi-sensor fusion, in: Intelligent Vehicles Symposium (IV), 2015 IEEE, IEEE, 2015, pp. 192–198.
[51] L. Xiao, B. Dai, D. Liu, D. Zhao, T. Wu, Monocular road detection using structured random forest, Int. J. Adv. Robot. Syst. 13 (2016) 101.
[52] L. Xiao, B. Dai, T. Wu, Y. Fang, Unstructured road segmentation method based on dictionary learning and sparse representation (in Chinese), J. Jilin Univ. 43 (2013) 384–388.
[53] W. Xiao, H. Liu, F. Sun, H. Liu, Likelihood confidence rating based multi-modal information fusion for robot fine operation, in: 2014 13th International Conference on Control Automation Robotics Vision (ICARCV), 2014, pp. 259–264, doi:10.1109/ICARCV.2014.7064316.
[54] R. Zhang, S.A. Candra, K. Vetter, A. Zakhor, Sensor fusion for semantic segmentation of urban scenes, in: Robotics and Automation (ICRA), 2015 IEEE International Conference on, IEEE, 2015, pp. 1850–1857.
[55] W. Zhu, J. Miao, J. Hu, L. Qing, Vehicle detection in driving simulation using extreme learning machine, Neurocomputing 128 (2014) 160–165. http://dx.doi.org/10.1016/j.neucom.2013.05.052.

Please cite this article as: L. Xiao et al., Hybrid conditional random field based camera-LIDAR fusion for road detection, Information Sciences (2017), http://dx.doi.org/10.1016/j.ins.2017.04.048