2015 IEEE Intelligent Vehicles Symposium (IV)
June 28 - July 1, 2015. COEX, Seoul, Korea

CRF based Road Detection with Multi-Sensor Fusion

Liang Xiao, Bin Dai, Daxue Liu, Tingbo Hu and Tao Wu

Abstract— **In this paper, we propose to fuse the LIDAR and monocular image in the framework of conditional random field to detect the road robustly in challenging scenarios. LIDAR points are aligned with pixels in image by cross calibration. Then boosted decision tree based classifiers are trained for image and point cloud respectively. The scores of the two kinds of classifiers are treated as the unary potentials of the corresponding pixel nodes of the random field. The fused conditional random field can be solved efficiently with graph cut. Extensive experiments tested on KITTI-Road benchmark show that our method reaches the state-of-the-art.**

I. INTRODUCTION

Road detection is a fundamental task in autonomous driving and has been studied for decades. On well marked roads, road detection can been done by detecting the lane marking which is much easier. However, general road detection is much more challenging due to the drastic change of road scenes, illumination, weather condition and the clutter of background.

In the literature, various systems based on different sensors have been developed for road detection. The mostly used is monocular image, which is cheap in costs but rich in information. Other sensor configurations like stereo vision and laser range finders are also widely used. Different kind of sensors have their own features. Monocular vision can get rich information of the field of view but it can be seriously affected by illumination. Especially, monocular vision can not get 3D information of the scene. Stereo vision can recover the 3D structure of the scene but dense stereo reconstruction is time consuming and the recovered 3D information can be noisome, especially with the increasing of distances. Active sensors like LIDARs(LIght Detector And Ranging) can measure the distance accurately, but the points lack information of texture and color. Furthermore, the points are always sparse, even in recently developed high definition LIDARs, like Velodyne HDL-64E.

Given the properties of the sensors, algorithms based on certain kind of sensor have their features too. Monocular vision based road detection usually learn classifier on labelled data or assume the bottom part of the image to be road and then group the pixels into road and background. This kind of method depend on the pixel feature extracted and the classifier employed. When the pixel feature are not discriminative between the road and the non-road areas, for example, the similarly paved side-walks, this kind of methods may fail. Besides, these methods can be sensitive to illumination.

Stereo vision based approaches first get dense disparity map through stereo matching. Then the image enhanced with depth can be utilized to detect the road. However, it is a contradiction between the computational cost and recover accuracy. Additionally, like monocular vision, stereo can also be seriously affected by illumination.

High definition LIDARs based road detection use the accurate 3D information acquired from the LIDAR points to analysis the 3D structure of the scene and take the flat area without obstacles as road. This kind of methods use only the sparse 3D points while the color and texture information are lacked to distinguish the non-road areas that have little or no differences in height.

Since each modal of sensor has its weakness, multi-modal sensor fusion can be a straightforward solution to fill the gap. In the paper, we propose to fuse the information of LIDAR and monocular image to get reliable road detection. The Velodyne HDL-64E LIDAR and a monocular camera are cross calibrated to get aligned. Then sensor fusion are integrated into the framework of conditional random field, in which the cues from image and LIDAR are processed and then jointly optimized with graph cuts. With the information of image and LIDAR used simultaneously, we can improve the performance. We tested our algorithm on the KITTI-Road dataset[1] and the results show that the proposed method reaches the state-of-the-art.

The remainder of this paper is organised as follows: First, the related works in road detection are briefly introduced in section II. The Section III introduces the methodology in detail and section IV shows the experimental results. Finally we draw the conclusion in Section V.

II. RELATED WORKS

Road detection has been a active research area since decades ago. A variety of systems has been developed to detect the road in some kind of environment. They used different sensors to acquire the information of the environment, such as, monocular vision, stereo vision , laser range finders and fusion of some of them.

For monocular vision based road detection, most approaches are learning based. These approaches often take the assumption of the lower part of the image being road and learn a classifier to discriminate the road and non-road pixels or patches. If offline data are available, the classifier can also be trained on the offline datasets after manually labelled. For learning based approaches, the feature and classifier are crucial for the performance. Usually, low level cues like

*This work was supported by National Scientific Foundation of China with Grant NO: 61375050

The authors are with the College of Mechatronic Engineering and Automation, National University of Defense Technology, Changsha, Hunan, P.R.China. {xiaoliang.cs, bindai.cs}@gmail.com

color, texture or combination of them are employed as the feature. In the literature, different color spaces have been investigated, such as HSI[2], normalized R-G[3], illumination invariant intrinsic image[4], [5] and linear combination of color planes[6], [7]. Texture are also frequently used, such as [8] and [9]. Other kind of features or variants are also used, such as slow feature analysis[10] and spatial ray feature[11]. For classifiers, various machine learning methods have been applied, such as, Gaussian mixture model[12], support vector machines(SVM)[13], boosting[10] and neural networks[14].

As pixel or patch based classifiers do not take the contextual information into consideration, the neighbouring units are classified independently and may generate unreasonable results. A commonly used solution to the problem is Random Field, which is very popular in semantic labelling[15]. Conditional Random Fields(CRFs) have been applied by many road detection algorithms. These algorithms often employ the framework of TextonBoost[15] which takes the pixel classifier output as the unary potential and takes the contrast dependent smooth prior as the pairwise potential. They are different in pixel feature and classifier. In [16], Alvarez proposed to use multi-scale convolution neural network as the classifier. Guo[17] proposed to combine random fields with unsupervised learning. There also some variants which take other prior into consideration. For example, He[18] proposed to use road shape prior to constrain the shape of the output.

For stereo vision based approaches, the disparity map is acquired through stereo matching. Then the disparity map can be analysed to get the free space and the road. In the literature, the V-disparity[19] were frequently used. In [20], a stochastic occupancy grid and dynamic programming based approach was proposed. These methods need dense stereo matching which is time consuming and the error increases with the distance.

Recently, with the advent of high definition LIDARs, several LIDAR based road detection algorithms have been developed. This kind of methods use the accurate 3D location of the LIDAR points to analyse the scene and take the flat area as the road. Thrun[21] presented a min-max elevation map based method which was widely used in DARPA Urban Challenge 2007. Moosmann[22] proposed to segment the LIDAR data using a local convexity criterion. Chen proposed to use Gaussian Process Regression in the polar grid map to segment the road area[23]. These methods can get good results when the difference of height between the road area and non-road area is significant. But when the difference is not salient in height but resides in other kind of features, say, color and texture, LIDAR based algorithms may fail.

To overcome the weakness of the single modal of sensor, multi-modal sensor fusion has been used. Vitor[24] developed a road detection system based on fusion of 2D and 3D vision. The 2D image are used to generate clusters by watershed. And stereo vision were used for free space analysis. Then the 2D and 3D features of each cluster are fused and fed into a Artificial Neural Network(ANN) to get classified. Shinzato[25] proposed to project the LIDAR points into the image plane and then triangulate the projected points to get a graph. Then the local spatial relationship between points is analysed to get the obstacles and the road is estimated by multiple free space detection. The method used the sensor fusion in a primitive way, *i.e.*, the method did not really use the information of image, it only used the cross calibration parameter to get the LIDAR points projected into the image plane. In contrast, sensor fusion used in [26] is more natural. The LIDAR points are firstly used to estimate the road plane. Then the road points are projected into the image and taken as seeds for learning a Gaussian model. The method differ from ours in two aspects. First, [26] use the 3D information of LIDAR points for road plane fitting. However, in our method, the LIDAR points are used in a data-driven way; it is totally learning based and no road model are assumed. Second, our method fuse the cues of LIDAR and image in the framework of conditional random field and they are jointly optimized to get a balanced result. While in [26], LIDAR cues were discarded after generating the seeds so the error in road plane fitting propagated to the following steps.

III. FUSED CRF FOR ROAD DETECTION

A. CRF for Road Detection

Since Shotton's seminal work[15], conditional random fields have been popular in semantic labelling. For road detection, a two-class(*road* and *background*) labelling problem can be adopted. In pixel-based CRFs, each pixel $i \in V = \{1, 2, \cdots N\}$ is modelled with a discrete random variable X_i which take one label from $\{road, background(bg)\}$. A neighbourhood system is defined with a set of N_i, for each $i \in V$, N_i defines all the neighbours of X_i, usually a locally 4-connected or 8-connected neighbourhood system are adopted.

Within this formulation, the posterior probability of the overall labelling given the observed image y can be expressed as

$$P(x|y) = \frac{1}{Z} \exp\left(\sum_{c \in C} -\psi_c(x_c) \right)$$

in which C is the set of all cliques and ψ_c is the potential function of clique c. Then the most probable labelling of x can be written as:

$$x^* = \underset{x}{argmin} - \log P(x|y) = \underset{x}{argmin} \log Z + E(x) = \underset{x}{argmin} E(x)$$

in which $E(x)$ is the corresponding Gibbs energy:

$$E(x) = \sum_{c \in C} \psi_c(x_c)$$

In the mostly used pairwise CRFs, only the unary potential and pairwise potential are considered, then the Gibbs energy can be written as:

$$E(x) = \sum_{x_i \in X} \psi_i(x_i) + \sum_{x_i, x_j; j \in N_i} \psi_{ij}(x_i, x_j). \qquad (1)$$

Just as TextonBoost[15] does, the unary potential $\psi_i(x_i)$ take the negative log-likelihood predicted by the learned pixel classifier: $\psi_i(x_i) = -\log p(x_i)$. And the pairwise potential $\psi_{ij}(x_i, x_j)$ penalized the neighbouring pixels which take different labels. As is done in [27], we can take the form of pairwise potential as:

$$\psi_{ij}(x_i, x_j) = \begin{cases} 0, & \text{if } x_i = x_j \\ \lambda \cdot \frac{1}{\text{dist}(i,j)} \cdot \exp\left(-\frac{\|I_i - I_j\|_2^2}{2\beta}\right), & \text{otherwise.} \end{cases} \quad (2)$$

in which I_i is the vector of the RGB values of the pixel x_i, $\text{dist}(i, j)$ is the Euclidean distance between the neighbouring pixel node i and j, β equals the expectation of the square contrast over the whole image and λ is the parameter control the trade-off between the strength of the unary potential and the contrast-dependent smooth prior.

B. *CRF for Road Detection with Sensor Fusion*

Although the aforementioned basic CRF yields fairly good results, it can suffer from the following problems: the image contains no 3D information and the only information we can use is the pixel appearance, however, the road areas in some scenarios can be very ambiguous in pixel appearance with the non-road areas such as the side-walk. As is shown in Fig. 1, the side-walk look extremely similar with the road area in pixel and the side-walk is liable to be mistakenly recognized as road without other cues.

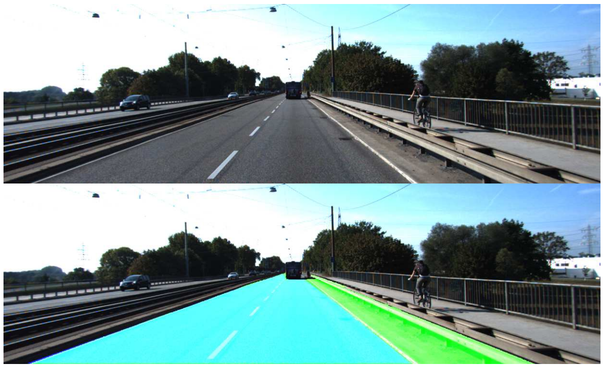

Fig. 1. The source image and the road predicted by basic pixel based pairwise CRF. The green channel denotes the predicted road area and the blue channel denotes the groundtruth; The main error is due to the similar looking of side-walk with the road.

When the autonomous vehicle is equipped with 3D LIDAR at the same time, the LIDAR points can offer the supplementary accurate 3D information that is lacked in monocular image. For non-road areas that is ambiguous in pixel appearance but different in height, it is easy to discriminate with LIDAR. So with multi-modal sensor fusion, we can improve the road detection substantially. In sensor fusion, temporal and spatial synchronization is very important. As is introduced in [28], LIDAR and camera are temporally synchronized by hardware trigger. With cross calibration[29], the LIDAR data and the image can be aligned. As is done in [25], we simply project the LIDAR points into the image plane and discard the points out of the field of view of the image. Fig. 2 shows a example of LIDAR and image alignment. From the tree trunks in the image, we can see that LIDAR and image are well aligned.

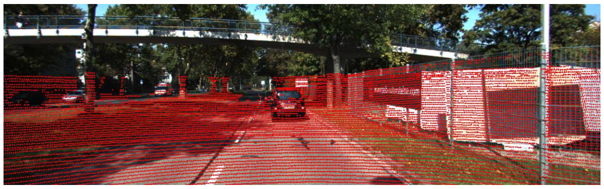

Fig. 2. Demo of LIDAR and image fusion, note the well aligned tree trunks

With the aligned image and LIDAR data, we can fuse the additional LIDAR information in the CRF. The LIDAR points projected into the image plane offer a strong cue for the corresponding pixels. We train a point cloud classifier that distinguish the road points with the other points. Then for the pixel nodes projected from LIDAR points, we can get another probabilistic prediction of being road or background. This can be seen as another kind of potential terms in the energy function to be minimized. Now the fused CRF have the following form of Gibbs energy:

$$E(x) = \sum_{x_i \in X} \psi_i(x_i) + \sum_{x_i, x_j; j \in N_i} \psi_{ij}(x_i, x_j) + \sum_{x_i \in X} \phi_i(x_i) \quad (3)$$

in which the additional energy term $\phi_i(x_i)$ represents the cue observed with the LIDAR data. Note that only a part of the pixels is aligned with LIDAR points. For these pixels aligned with LIDAR points(denoted as set P), the potential terms take the form of

$$\phi_i(x_i) = -\gamma \cdot \log p'(x_i), i \in P \quad (4)$$

in which γ is the parameter control the tradeoff between the three energy terms in (3), p' are the probabilistic output of LIDAR classifier(introduced in the following subsection). While for the rest pixels, we get no additional information stemming from LIDAR scans. We simply take the probability of being road and background predicted by LIDAR equally as $1/2$ and therefore take the potential $\phi_i(x_i)$ as $-log(1/2)$.

C. *Classifiers Training*

As we use the output of the classifier as the unary potential, we need first label some data acquire in typical scenarios. While labelling the image can be rather simple, labelling the point cloud is much more labour-consuming. So we label the images and then transfer the label information to the corresponding point clouds with the cross-calibration parameters.

After labelling, we can train classifiers. In this paper, we choose to use the boosted decision tree as the classifier for both image and point cloud. Using the decision tree of depth d as the weak classifiers T_i, we can get a strong classifier T by running adaBoost for K rounds:

$$T = \sum_{i=1}^{K} \alpha_i T_i$$

Then the class scores(sc) of each pixel or LIDAR point z being road*(1)* or background*(0)* can be denoted as,

$$sc(road|z) = \sum_{i=1}^{K} \alpha_i \delta(T_i(z))$$

$$sc(bg|z) = \sum_{i=1}^{K} \alpha_i (1 - \delta(T_i(z)))$$

in which, δ is the Kronecker delta function and $T_i(z)$ is the prediction of the *ith* weak classifier.

Then the class score can be transformed to a normalized probability by

$$p(road|z) = \frac{sc(road|z)}{sc(road|z) + sc(bg|z)},$$

$$p(bg|z) = \frac{sc(bg|z)}{sc(road|z) + sc(bg|z)}.$$

D. Feature Extraction

1) Image Features: For image based features, we extracted pixel-wise *Texture*, *Dense HOG* and *Color* as the feature. We also include the *location* cues in the feature.

- *Filter Bank.* The images were converted to the CIE *Lab* color space and then the filter bank were applied to the gray scale image or each channel of the CIE *Lab* image. Concretely, Gaussian filter were applied to each channel while the horizon and vertical Gaussian Derivative filters and the Laplacian of Gaussian filter were applied to the gray scale image. Therefore, for a given scale σ , we get a 6-dimensional feature for each pixel. In this paper, three scales were employed so we got a 18-dimensional filter bank response for each pixel.
- *LBP.* 4-connected neighbouring local binary pattern feature is extracted to describe the local texture additionally.
- *Dense HOG.* Dense Histogram of Oriented Gradients are calculated for 9 directions.
- *Color and location.* Instead of making the assumption of the lower part of the images being road, we take the normalized location(*w.r.t* the image size) of the pixel as part of the feature. This make the location cues effectively utilized while in a more reasonable way than the simple assumption. In addition, we also included the RGB channels in the feature.

So finally, we got a 36-dimensional feature vector for each pixel in the image. The feature vectors are then fed into the classifier to get the probability of being road or background for each pixel.

2) Point Cloud Feature: For point cloud feature, we use the simple geometric feature, including the normalized 3D location(*w.r.t* the Euclidean distance) and the direction of the local normal vector Note that we take the simple feature because it is fast to compute and it suffice to demonstrate the rationale of the algorithm. See Fig. 3 for the results of point cloud classified with the aforementioned feature. We can get better results with more powerful point cloud features like spin image and shape context. See [30] for a review.

E. Graph cuts

To get the most probable labelling we need to solve the energy minimization problem of (3). That kind of energy minimization problem is popular in computer vision and can be solved efficiently with graph cut. In this paper, we employ the open source library Darwin[31] to solve the energy minimization problem of fused CRF.

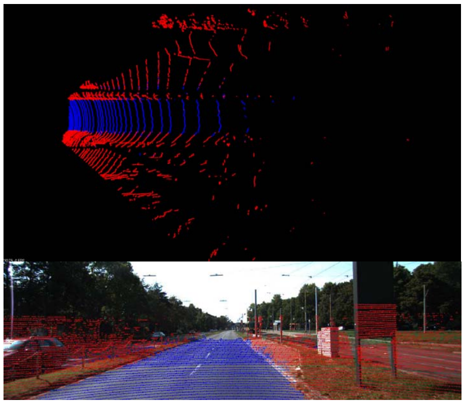

Fig. 3. Output of point cloud classifier, blue point are classified as road

IV. EXPERIMENTS

A. Datasets

To validate the effectiveness of the proposed method, we tested the proposed algorithm on the KITTI Benchmark[28]. The KITTI-Road dataset[1] contain synchronized images and LIDAR data with calibration parameters. The road and lane estimation benchmark consists of 289 training and 290 test images. It contains three different categories of road scenes: UU(urban unmarked), UM(urban marked), and UMM(urban multiple marked lanes). In this paper, we only deal with road detection, the lane information is not considered here. The dataset offer groundtruth for training data and online evaluation of testing data is offered on the website. For evaluation, the dataset offered pixel-based evaluation and behaviour-based evaluation. A set of metrics including precision(**PRE**), recall(**REC**), maximum F1-measure(**MaxF**) ,average precision(**AP**), false positive rate(**FPR**) and false negative rate(**FNR**) are used for evaluation.

B. Experiment setting

The detailed parameter settings are as follows, for pixel and point cloud classifier training, we take decision tree with depth $d = 4$ as the weak classifier. And we run AdaBoost for $K = 50$ rounds. For conditional random field, we adopt the locally 8-connected neighbourhood system. The two parameters λ and γ in (2) and (4) act important roles in the proposed algorithm and are to be analysed in detail in the following subsection. The experiments were tested on a standard PC with 8GB of RAM and a dual-core *Intel Core i5-3230M* CPU clocked at 2.6GHz. The algorithm was implemented with C++ under Ubuntu 12.04. As we take each pixel as a random variable, for images from KITTI-Road dataset(with resolution of about 1240×375), the average time consuming is about 2 seconds.

C. Evaluation

In this part, we show the qualitative and quantitative evaluation of the proposed algorithm. Because the groundtruths of the testing set are not offered, we use only the 289 training images for the first validation experiment. The training data of the each of the three branches of the dataset are randomly divided into two equal parts. One used for training and the other used for testing. Take the UM sub-dataset for example, we divided the original training set with 95 labelled images into two sets, one with 47 images for training and the other for testing(*called new UM training set and new UM testing set respectively later on*). Then we train the pixel and point cloud classifier on the new training sets and test on the new testing sets.

The parameters λ and γ in equation (3) tradeoff between the pixel classifier, point cloud classifier and the smooth prior. To analysis the impact of the parameters, we tested different setting of λ and γ on the new UM testing set. We accumulated the number of the mistakenly classified pixels over the testing images for different parameters configuration. Fig. 4 illustrates the variation of the error with respect to the parameters. From the figure, we can see that when the parameter λ or γ is very small, the error is very big. And when the parameters increases, the error drop dramatically. But when the parameters grow too big, the error increases slowly. This can help us choose the suitable parameters.

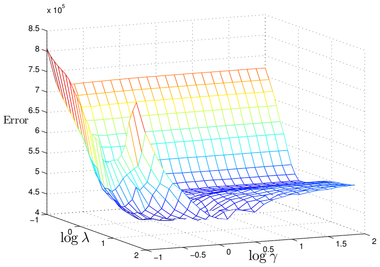

Fig. 4. Accumulated error with respect to different setting of λ and γ

To show the priority of the proposed algorithm(*fused CRF*), we compared the results with the basic pixel based pairwise CRF(*pairwise CRF*). Additionally, we take the output of pixel classifier(*Boosting*) as the baseline. For pairwise CRF and fused CRF, we tune the best parameters λ and γ for UM, UMM and UU subsets respectively on the three new testing sets. Although the best parameters for the three subsets are slightly different, setting $\lambda = 10$ and $\gamma = 45$ yields fairly good results for all. Fig. 5 show a example of the results. We can see from the figure that some area that are very ambiguous in pixel appearance are easy to distinguish with 3D information acquired form LIDAR. So with sensor fusion, we can enhance the performance. Now we evaluate the results quantitatively in the perspective view. Table I, Table II and Table III show the quantitative evaluation[%] on the new UM, UMM and UU testing sets. From the tables, we can see that with the local smooth prior, pairwise CRF get better performance over the pixel based classifier. And with the information of LIDAR fused into the CRF, we can boost the performance further.

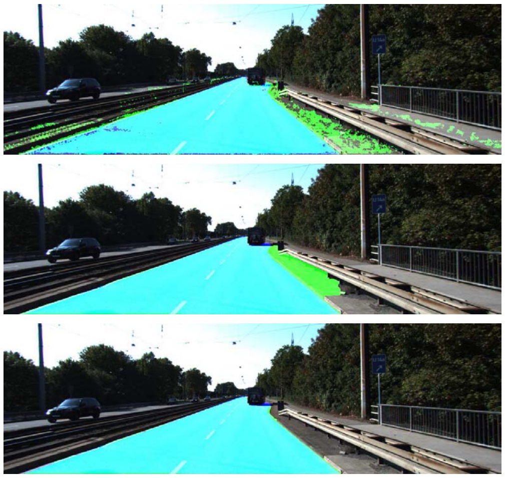

Fig. 5. Illustration of the priority of the proposed algorithm, the first row shows the result of basic pixel classifier, the middle row shows the results of pixel based CRF and the third row shows the results of the proposed fused CRF. Predicted road areas are overlapped with green, and groundtruths are overlapped with blue.

TABLE I
COMPARISON ON UM(PERSPECTIVE VIEW)

	Boosting	Pairwise CRF	Fused CRF(Ours)
MaxF	88.62	91.24	**94.17**
AP	76.95	80.83	**85.63**
PRE	83.00	87.27	**92.56**
REC	95.06	95.59	**95.84**
FPR	3.82	2.71	**1.51**
FNR	4.94	4.74	**4.16**

Then we use all the training images for training and evaluate the results in the bird eye view(BEV) on the website[1]. The parameters λ and γ are setted as the same as the last experiments. Fig. 6 shows some results in BEV. The first column shows two images from UM set, the second column shows two images from UMM set and the third column shows two images from UU set. Note that red denotes false negatives, blue areas correspond to false positives and green represents true positives.

We compared our algorithm with the recently developed ones, including, CN[7], SPRAY[32], ProbBoost[33], RES3D-Velo[25], HistonBoost[34], BM[5], ARSL-AMI[35], SPlane+BL[36], and the baseline(BL)[1]. The evaluation in BEV on UM, UMM, UU sets and the average results on the all sets(URBAN_ROAD) are shown in table IV, V, VI, and VII. From the results showed in the tables, we can see that our algorithm get state-of-the-art results on the UM set and UU set. But the results on the UMM set is less competitive than those on UM and UU sets, that may be due to the images in UMM set contain little high rising non-road objects and the 3D information of LIDAR are less helpful. And in average, the proposed algorithm gets the best **MaxF**. Note that the average precision(**AP**) of the proposed algorithm is quite low; that is because the average precision is a description of Precision-Recall curve with different

[1] http://www.cvlibs.net/datasets/kitti/eval_road.php

TABLE II

COMPARISON ON UMM(PERSPECTIVE VIEW)

	Boosting	Pairwise CRF	Fused CRF(Ours)
MaxF	91.29	92.39	**93.50**
AP	82.17	83.32	**85.79**
PRE	87.95	89.22	**91.94**
REC	94.89	**95.79**	95.11
FPR	4.18	3.72	**2.68**
FNR	5.11	**4.21**	4.89

TABLE III

COMPARISON ON UU(PERSPECTIVE VIEW)

	Boosting	Pairwise CRF	Fused CRF(Ours)
MaxF	82.27	86.32	**90.35**
AP	68.22	75.60	**81.78**
PRE	73.62	81.73	**88.52**
REC	**93.22**	91.44	92.25
FPR	5.57	3.41	**1.99**
FNR	**6.78**	8.56	7.75

thresholds to classify the confidence maps. But the proposed algorithm generate binary output, so it is not suitable to be evaluated in that metric. This is also argued in [5].

V. CONCLUSIONS AND FUTURE WORKS

In this paper, a LIDAR and image fusion based road detection algorithm is proposed. We formulate the road detection as a two class pixel labelling problem and solve it within conditional random field framework. We learn boosted tree classifiers for pixels and point cloud respectively. During testing, the output of pixel and point cloud classifiers are employed as the potential terms of the corresponding pixels and the pixels located at the projection of LIDAR points into the image. Then the optimization of the fused CRF is solved with graph cut. The experiments tested on KITTI-Road Benchmark show the priority of the proposed fused CRF over the basic pixel base pairwise CRF. And the online evaluation show that the proposed algorithm reaches the state-of-the-art.

The proposed method uses only the most simple features for point cloud, we are considering to employ more sophisticated features to get better performance. Besides, in this paper, we build the random field model on the pixel lattice and it is time consuming due to the large number of the random variables. There are two straightforward ways to speed up. First, group the image into superpixels or patches and use the LIDAR points that projected into the units to reduce the uncertainty. The other way is to take the label of each LIDAR point as a random variable and extend the points with the features of the aligned pixels. Besides, higher order CRF or fully connected CRF can be employed to get better results.

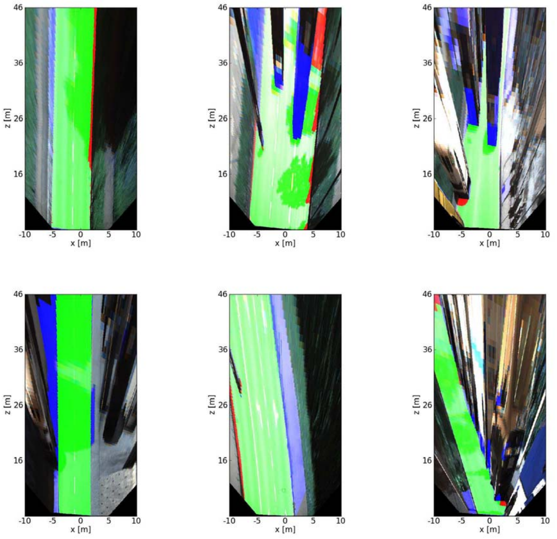

Fig. 6. Road detected in bird eye view. Here, red denotes false negatives, blue areas correspond to false positives and green represents true positives. Best viewed in color

TABLE IV

RESULTS OF ONLINE EVALUATION ON UM(BEV)

Algorithm	**MaxF**	***AP***	**PRE**	**REC**	**FPR**	**FNR**
CN	73.69	76.68	69.18	78.83	16.00	21.17
SPRAY	88.14	**91.24**	**88.60**	87.68	**5.14**	12.32
BL	82.24	85.30	79.44	85.24	10.05	14.76
ProbBoost	87.48	80.13	85.02	90.09	7.23	9.91
HistonBoost	83.68	72.79	82.01	85.42	8.54	14.58
BM	78.90	66.06	69.53	91.19	18.21	8.81
ARSL-AMI	71.97	61.04	78.03	66.79	8.57	33.21
SPlane+BL	85.23	88.66	83.43	87.12	7.89	12.88
RES3D-Velo	83.81	73.95	78.56	89.80	11.16	10.20
FusedCRF(Ours)	**89.55**	80.00	84.87	**94.78**	7.70	**5.22**

TABLE V

RESULTS OF ONLINE EVALUATION ON UMM(BEV)

Algorithm	**MaxF**	***AP***	**PRE**	**REC**	**FPR**	**FNR**
CN	86.21	84.40	82.85	89.86	20.45	10.14
SPRAY	89.69	**93.84**	**89.13**	90.25	**12.10**	9.75
BL	76.02	78.82	65.71	90.17	51.72	9.83
ProbBoost	**91.36**	84.92	88.18	94.78	13.97	5.22
HistonBoost	88.73	81.57	84.49	93.42	18.85	6.58
BM	89.41	80.61	83.43	**96.30**	21.02	**3.70**
ARSL-AMI	89.56	82.82	85.87	93.59	16.93	6.41
SPlane+BL	82.04	85.56	75.11	90.39	32.93	9.61
RES3D-Velo	90.60	85.38	85.96	95.78	17.20	4.22
FusedCRF(Ours)	89.51	83.53	86.64	92.58	15.69	7.42

TABLE VI

RESULTS OF ONLINE EVALUATION ON UU(BEV)

Algorithm	**MaxF**	***AP***	**PRE**	**REC**	**FPR**	**FNR**
CN	72.25	66.61	71.96	72.54	9.21	27.46
SPRAY	82.71	**87.19**	82.16	83.26	5.89	16.74
BL	69.50	73.87	65.87	73.56	12.42	26.44
ProbBoost	80.76	68.70	**85.25**	76.72	4.33	23.28
HistonBoost	74.19	63.01	77.43	71.22	6.77	28.78
BM	78.43	62.46	70.87	87.80	11.76	12.20
ARSL-AMI	70.33	61.97	83.33	60.84	**3.97**	39.16
SPlane+BL	74.02	79.61	65.15	85.68	14.93	14.32
RES3D-Velo	83.63	72.58	77.38	90.97	8.67	9.03
FusedCRF(Ours)	**84.49**	72.35	77.13	**93.40**	9.02	**6.60**

TABLE VII

Averagre Results on Urban_Road(BEV)

Algorithm	**MaxF**	***AP***	**PRE**	**REC**	**FPR**	**FNR**
CN	79.02	78.80	76.64	81.55	13.69	18.45
SPRAY	87.09	**91.12**	87.10	87.08	**7.10**	12.92
BL	75.80	79.85	69.31	83.63	20.40	16.37
ProbBoost	87.78	77.30	**86.59**	89.01	7.60	10.99
HistonBoost	83.92	73.75	82.24	85.66	10.19	14.34
BM	83.47	72.23	75.90	92.72	16.22	7.28
ARSL-AMI	80.36	70.23	83.24	77.67	8.61	22.33
SPlane+BL	79.63	83.90	72.59	88.17	18.34	11.83
RES3D-Velo	86.58	78.34	82.63	90.92	10.53	9.08
FusedCRF(Ours)	**88.25**	79.24	83.62	**93.44**	10.08	**6.56**

References

[1] J. Fritsch, T. Kuehnl, and A. Geiger, “A new performance measure and evaluation benchmark for road detection algorithms,” in *International Conference on Intelligent Transportation Systems (ITSC)*, 2013.

[2] M. A. Sotelo, F. J. Rodriguez, and L. Magdalena, “Virtuous: Vision-based road transportation for unmanned operation on urban-like scenarios,” *Intelligent Transportation Systems, IEEE Transactions on*, vol. 5, no. 2, pp. 69–83, 2004.

[3] C. Tan, T. Hong, T. Chang, and M. Shneier, “Color model-based real-time learning for road following,” in *Intelligent Transportation Systems Conference, 2006. ITSC ’06. IEEE*, Sept 2006, pp. 939–944.

[4] J. Alvarez and A. M. Lopez, “Road detection based on illuminant invariance,” *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 1, pp. 184–193, 2011.

[5] B. Wang, V. Fremont, and S. Rodriguez, “Color-based road detection and its evaluation on the kitti road benchmark,” in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, June 2014, pp. 31–36.

[6] J. M. Alvarez, M. Salzmann, and N. Barnes, “Learning appearance models for road detection,” in *Intelligent Vehicles Symposium (IEEE IV)*. IEEE, 2013.

[7] J. M. Alvarez, T. Gevers, Y. LeCun, and A. M. Lopez, “Road scene segmentation from a single image,” in *European Conference on Computer Vision, ECCV*, 2012.

[8] C. Rasmussen, “Grouping dominant orientations for ill-structured road following,” in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1. IEEE, 2004, pp. I–470.

[9] H. Kong, J.-Y. Audibert, and J. Ponce, “Vanishing point detection for road detection,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 96–103.

[10] T. Kuhnl, F. Kummert, and J. Fritsch, “Monocular road segmentation using slow feature analysis,” in *Intelligent Vehicles Symposium (IV), 2011 IEEE*, June 2011, pp. 800–806.

[11] J. Fritsch, T. Kuehnl, and F. Kummert, “Monocular road terrain detection by combining visual and spatial information,” *Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1586–1596, 2014.

[12] H. Dahlkamp, A. Kaehler, D. Stavens, S. Thrun, and G. R. Bradski, “Self-supervised monocular road detection in desert terrain,” in *Robot. Sci. Syst. Conf. (RSS)*, 2006.

[13] Y. Alon, A. Ferencz, and 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*, ser. CVPR ’06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 689–696.

[14] P. Y. Shinzato, V. G. Jr, F. S. Osorio, and D. F. Wolf, “Fast visual road recognition and horizon detection using multiple artificial neural networks,” in *Intelligent Vehicles Symposium Proceedings, 2012 IEEE*, June 2012, pp. 1090–1095.

[15] J. Shotton, J. Winn, C. Rother, and 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.

[16] J. M. Alvarez, Y. LeCun, T. Gevers, and 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.

[17] C. Guo, S. Mita, and D. McAllester, “Robust road detection and tracking in challenging scenarios based on markov random fields with unsupervised learning,” vol. 13, no. 3, pp. 1338–1354, Sep 2012.

[18] Z. He, T. Wu, Z. Xiao, and H. He, “Robust road detection from a single image using road shape prior,” in *Image Processing (ICIP), 2013 20th IEEE International Conference on*, Sept 2013, pp. 2757–2761.

[19] R. Labayrade, D. Aubert, and J.-P. Tarel, “Real time obstacle detection in stereovision on non flat road geometry through ”v-disparity” representation,” in *Intelligent Vehicle Symposium, 2002. IEEE*, vol. 2, June 2002, pp. 646–651.

[20] H. Badino, U. Franke, and R. Mester, “Free space computation using stochastic occupancy grids and dynamic programming,” in *Workshop on Dynamical Vision, ICCV, Rio de Janeiro, Brazil*, vol. 20, 2007.

[21] 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.

[22] F. Moosmann, O. Pink, and C. Stiller, “Segmentation of 3d lidar data in non-flat urban environments using a local convexity criterion,” in *Intelligent Vehicles Symposium, 2009 IEEE*. IEEE, 2009, pp. 215–220.

[23] T. Chen, B. Dai, R. Wang, and D. Liu, “Gaussian-process-based real-time ground segmentation for autonomous land vehicles,” *Journal of Intelligent & Robotic Systems (JINT)*, vol. 76, pp. 563–582, Sep 2013.

[24] G. B. Vitor, D. A. Lima, A. C. Victorino, and J. V. Ferreira, “A 2d/3d vision based approach applied to road detection in urban environments,” in *Intelligent Vehicles Symposium (IV), 2013 IEEE*, 2013, pp. 952–957.

[25] P. Y. Shinzato, D. F. Wolf, and C. Stiller, “Road terrain detection: Avoiding common obstacle detection assumptions using sensor fusion,” in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, June 2014, pp. 687–692.

[26] X. Hu, S. A. R. F., and A. Gepperth, “A multi-modal system for road detection and segmentation,” in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, June 2014, pp. 1365–1370.

[27] C. Rother, V. Kolmogorov, and A. Blake, “”GrabCut”: Interactive foreground extraction using iterated graph cuts,” in *ACM SIGGRAPH 2004 Papers*, ser. SIGGRAPH ’04. New York, NY, USA: ACM, 2004, pp. 309–314.

[28] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The KITTI dataset,” *The International Journal of Robotics Research*, vol. 32, pp. 1231–1237, 2013.

[29] A. G. F. Moosmann, O. Car, and B. Schuster, “Automatic camera and range sensor calibration using a single shot,” in *IEEE International Conference on Robotics and Automation, ICRA*, 2012, pp. 3936–3943.

[30] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan, “3d object recognition in cluttered scenes with local surface features: A survey,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 11, pp. 2270–2287, Nov 2014.

[31] S. Gould, “DARWIN: A framework for machine learning and computer vision research and development,” *Journal of Machine Learning Research (JMLR)*, vol. 13, pp. 3533–3537, Dec 2012.

[32] T. Kuehnl, F. Kummert, and J. Fritsch, “Spatial ray features for real-time ego-lane extraction,” in *Proc. IEEE Intelligent Transportation Systems*, 2012.

[33] G. B. Vitor, A. C. Victorino, and 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*, May 2014.

[34] G. Vitor, A. Victorino, and 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)*, June 2014, pp. 19–24.

[35] M. Passani, J. J. Yebes, and L. M. Bergasa, “CRF-based semantic labeling in miniaturized road scenes,” in *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, Oct 2014, pp. 1902–1903.

[36] N. Einecke and J. Eggert, “Block-matching stereo with relaxed fronto-parallel assumption,” in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, June 2014, pp. 700–705.