



# Comprehensive performance analysis of road detection algorithms using the common urban Kitti-road benchmark

Giovani Bernardes Vitor, Alessandro C. Victorino, Janito V. Ferreira

## ► To cite this version:

Giovani Bernardes Vitor, Alessandro C. Victorino, Janito V. Ferreira. Comprehensive performance analysis of road detection algorithms using the common urban Kitti-road benchmark. IEEE Intelligent Vehicle Symposium (IV 2014), Jun 2014, Dearborn, Michigan, United States. pp.19 - 24, 10.1109/IVS.2014.6856616 . hal-01089094

**HAL Id: hal-01089094**

**<https://hal.science/hal-01089094>**

Submitted on 1 Dec 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Comprehensive performance analysis of road detection algorithms using the common urban Kitti-road benchmark

Giovani B. Vitor<sup>1,2</sup>, Alessandro C. Victorino<sup>1</sup> and Janito V. Ferreira<sup>2</sup>

**Abstract**—The navigation of an autonomous vehicle is a highly complex task and the dynamic environment is used as a source for reasoning. Road detection is a major issue in autonomous systems and advanced driving assistance systems applied for inner-city. Uncertainty may arise in environments with unmarked or weakly marked roads or poor lightning conditions. Moreover, when a common benchmark is not used, it is hard to decide which approach performs better on the road detection problem. This paper introduces a comprehensive performance analysis of two road recognition approaches using the urban Kitti-road benchmark. The first approach makes the extraction of a feature set based on statistical measures of 2D and 3D information from each superpixel. An Artificial Neural Network is used to detect the road pattern. The second approach extracts the feature set based on a multi-normalized histogram of Textons and Disptons for each superpixel. This feature set is used as a source for a Joint Boosting algorithm to model the road pattern. The proposed work presents a detailed evaluation highlighting the pros and cons of each approach.

**Index Terms**—Road Recognition, Computer Vision, Artificial Neural Network, Joint Boosting, Texton Map, Dispton Map, Watershed Transform.

## I. INTRODUCTION

In recent years, several applications for control of autonomous vehicles and specialized Advanced Driving Assistance Systems (ADAS) were proposed [1][2]. In most cases machine vision is used as a main source of information for road detection thanks to the facility to extract measurements related to texture and color. Furthermore, several vehicles are already equipped with cameras and this large-scale adoption could be motivated by their low cost. Specifically for ADAS systems, machine vision has been used to detect road surfaces where lane keeping assistance systems have already been embedded with a limited capability to detect roads with smooth curvatures [3].

Several approaches related to vision-based road detection have been proposed, varying from mono to stereo vision sources. They aim at solving several challenging problems such as the continuously changing backgrounds in different environments (inner-city, highway, off-road), different road types (shape and color), the presence of different objects (signs, vehicles, pedestrian) and also different imaging conditions (variation of illumination and weather conditions).

In order to provide support on unmarked roads that are common in inner-city environments, general approaches overcome the assumption of lane marking discriminating

the overall road area with some appearance cues, spacial information or the combination of both. Several works emphasize the color feature in their models [4][5][6], while others use the texture feature [7][8]. Outcomes from the color feature may present poor results when there is high intra-class variability presents in the dynamic nature of the scenes. Texture is scale-dependent and is affected by the strong perspective in road image. Some approaches propose improvements in the model by merging color, texture and adding 3D information [9][10][11].

All presented approaches have been evaluated using different datasets and criterias. Therefore a fair performance comparison is prohibitive [12]. Following the approach proposed by Fritsch et al. [12], this paper contributes to produce a standard performance analysis using the challenging urban KITTI-road benchmark as a dataset. This work proposes an evaluation of two distinct approaches highlighting the pros and cons of each one. The first approach makes the extraction of a feature set based on statistical measures of 2D and 3D information for each superpixel. An Artificial Neural Network is used to detect the road pattern. The second approach extracts the feature set based in a multi-normalized histogram of Textons and Disptons for each superpixel. This feature set is used as a source for a Joint Boosting algorithm to model the road pattern.

The rest of the paper is organized as follows: Section II presents the Artificial Neural Network rendering the 2D image segmentation and 3D image processing to model the road pattern. Section III presents the multi-normalized histogram with Joint Boost algorithm, covering the Texton maps and the Dispton maps. Section IV presents the results of the two approaches using the space and metrics perspective. Some conclusions presenting what we have learned from the study and proposing future work are presented in Section V.

## II. ARTIFICIAL NEURAL NETWORK FOR ROAD RECOGNITION

This section describes the first algorithm for road recognition. The algorithm executes the 2D image segmentation and 3D image processing to compose a feature's descriptor based on the texture information and spacial information. This feature's descriptor is used by the Artificial Neural Network (ANN) to learn the road pattern.

### A. 2D Image Processing

The *2D Image Segmentation* algorithm performs a combination of filters with *Watershed Transform* to get the superpixel. This segmentation process uses an approach inspired

Authors are members of <sup>1</sup>Heudiasyc UMR CNRS 7253 Université de Technologie de Compiègne, <sup>2</sup>Universidade Estadual de Campinas (UNICAMP). Giovani B. Vitor holds a Ph.D. scholarship from CNPQ. Contact authors [giovani.bernardes-vitor@hds.utc.fr](mailto:giovani.bernardes-vitor@hds.utc.fr)

by the techniques used by the domain of morphological image processing [13], where three morphological filters are applied to reach the superpixel, namely *Morphological Gradient Adjusted*, the *AreaClose* and the *Hmin*. The first one is applied to obtain the highest frequency of the image, where an adjustment is made to improve the low-contrast of high frequency in shadow areas [14]. The *AreaClose* filter is employed to filter out connected components smaller than a given parameter  $\lambda$ , followed by the *Hmin* filter that is responsible for eliminating the local minimum given a second parameter  $h$ . The main goal of these filters is to give flexibility allowing the Watershed Transform to determine the segmentation result, i.e., ranging the parameter  $\lambda$  and  $h$ , it is possible to define a fine or a coarse segmentation result. This process follows the same approach proposed in [11] and its result can be seen in Figure 1a.

### B. 3D Image Processing

The *3D Image Processing* algorithm uses the rectified pair of stereo images to allow an easy pre-classification for the image data into drivable and non-drivable areas. The approach takes advantage of the Epipolar geometry [15]. First, the disparity map ( $I_{\Delta}$ ) of the stereo pair is built using the Sum of Absolute Differences (SAD) correlation algorithm. As described in [16], the world environment can be approximated by horizontal and vertical planes related to the camera. Based on this assumption, each disparity map ( $I_{\Delta}$ ) are then projected in the v-disparity map ( $I_{v\Delta}$ ) [9], where each line on  $I_{v\Delta}$  represent these planes. The next step is the execution of the Hough Transform with an auxiliary algorithm to detect the lines associated with the drivable area. In  $I_{v\Delta}$ , the continuous slope lines represent the drivable area, while the non-drivable areas are approximately represented by vertical lines. As a result, the lines from Hough transform are filtered out considering a given parameter of inclination performed by the auxiliary algorithm. It applies also the moving average technique along the time, to avoid the noisy data in the  $I_{v\Delta}$  derived from a wrong correlation in  $I_{\Delta}$ . Finally, the detected lines are reprojected into the image, as can be seen in Figure 1b, which generates an estimation of drivable and non-drivable area.

### C. Artificial Neural Network algorithm

The Artificial Neural Network (ANN) using a Multilayer Perceptron (MLP) is applied to learn the road pattern resulting in the final classification of the road recognition. First, the method constructs a feature's descriptor based on the superpixel generated from the 2D image segmentation. 3D and 2D information are generated for each superpixel. The first set of features consider the percentage of the drivable and non-drivable areas where it is based on the intersection of the resulting pre-classification of 3D image processing with the respective superpixel. The second set of features is based on statistical measures, as proposed by [17], like mean, probability, entropy and variance. These values were

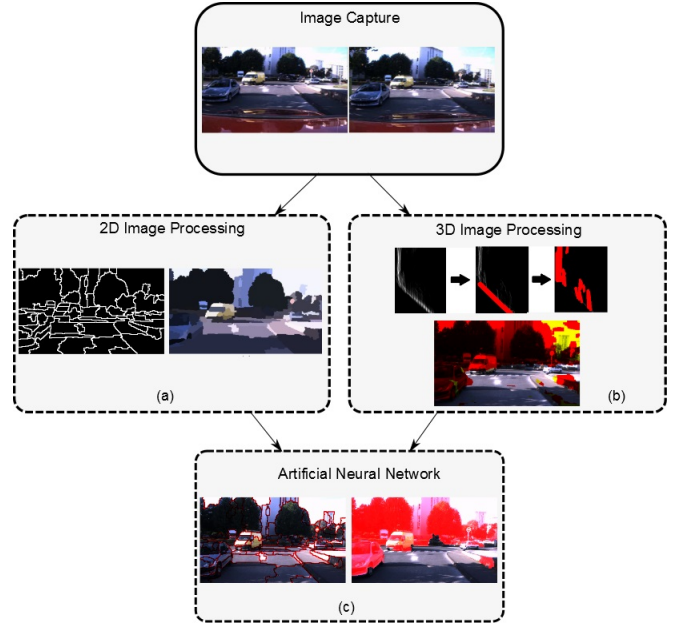


Fig. 1. The process overview for the first algorithm. (a) The 2D segmentation process to get the superpixel. (b) The 3D process to obtain an estimation of the drivable and non-drivable areas. (c) The final result applying the ANN, which the left image represents the lines of the superpixel and the right image has the classification of non-road surface (red) and road surface (original color).

calculated by the RGB<sup>1</sup> and HSV<sup>2</sup> color space values of each superpixel.

The network training is based on the Backpropagation technique and its structure was projected with three layers, namely input, output and hidden layers. The size of the input layer corresponds to the number of features extracted which was defined in the experiments with six neurons in its final version. The hidden layer was defined with 30 neurons, where all neurons use the sigmoid activation function. Finally, the output layer has two neurons to classify the feature's descriptor as road surface or non-road surface, which response vary from 0 to 1. Figure 1 presents all processes performed by this first approach.

## III. MULTI-HISTOGRAM-BASED JOINT BOOSTING FOR ROAD RECOGNITION

In this section is given a brief understanding of the second algorithm for road recognition. The method merges 2D and 3D information, resulting in a multi-normalized histogram that model the appearance, shape and context using a Joint Boost algorithm. Differently from the first one, this approach creates a form to represent the 3D information. It encapsulates the 3D information on *Disptons*<sup>3</sup> to compose jointly with the *Textons*<sup>4</sup>, a different set of features, to better represent the road class given by the complex environment.

<sup>1</sup>Abbreviation for red, green and blue color space.

<sup>2</sup>Abbreviation for hue, saturation and value color space.

<sup>3</sup>Dispton means the name of a proposed technique that encapsulates 3D information from disparity map in the called Dispton maps

<sup>4</sup>Textons are the result of a technique that performs a clustering based in a given feature set, extracted from texture image.

A discriminative model for road class is learned based in the multi-normalized histogram performed on superpixel, shown in section II-A, and classified using a Joint Boosting classifier.

#### A. Texton Maps

The *Textons* have been proven to be effective to discriminate between some classes of similar textures [18]. The method applies the textonization process based on the extraction of various features to get the *texton maps* [19]. This process of textonization construct a dictionary of *Textons* by using an unsupervised classifier. The unsupervised classifier is given by the K-Means algorithm where the *Dictionary of Textons*  $t_i \in D = \{t_1, \dots, t_K\}$  is defined by clusters generated according to the given feature. At the end, a particular *texton*  $t_i$  is associated to each pixel, providing a mapping of *Textons*  $T \in \mathbb{N}^2$  denoted by *texton map*.

The *textonization* is performed on a set of feature's descriptor, as can be seen in [18]. The set include 17-dimensional filter bank, 3-dimensional CIELAB color, 81-dimensional histograms of oriented gradient [20] and 2-dimensional normalized pixel location. All feature's descriptor are whitened (to give zero mean and unit covariance) to learn the dictionaries of textons in which their configuration were setted to 400 clusters, 128 clusters, 150 clusters and 144 clusters, respectively.

#### B. Dispton Maps

Based on Texton maps, Dispton maps are the outcomes of a technique that builds two additional dictionaries over 3D information from Stereo Vision. It aims at creating meaningful clusters based on Disparity maps ( $I_\Delta$ ). This technique is called by *Disptonization*. It encapsulates 3D information provided by the U-Disparity and V-Disparity in a dictionary of Dispton, taking then the *Dispton Maps* from  $I_\Delta$ . The summarized Disptonization algorithm can be seen in (1).

---

##### Algorithm 1 Disptonization algorithm:

---

- 1: Process  $I_u\Delta$  and  $I_v\Delta$  from  $I_\Delta$ ;
  - 2: Apply the *Hmin* filter on  $I_u\Delta$  and  $I_v\Delta$ ;
  - 3: Binarize and obtain the line segments by *Hough Transf.*;
  - 4: Determine the *U-Dispton* dictionary;
  - 5: Determine the *V-Dispton* dictionary;
  - 6: Generate the *UV-Dispton maps*;
- 

The disptonization process generates two features called V-Dispton map and U-Dispton Map, containing 4 clusters and  $N$  clusters respectively. The dispton maps can be seen in Figure 2.

#### C. Multi-Normalized-Histogram Joint Boosting algorithm

The multi-normalized-histogram joint boosting algorithm abbreviated by *HistonBoost*, is built to model the road class, using an adapted version of the joint boosting algorithm [21]. This algorithm is derived from *TextonBoost* [22], which iteratively builds a *strong classifier* as a sum of *week*

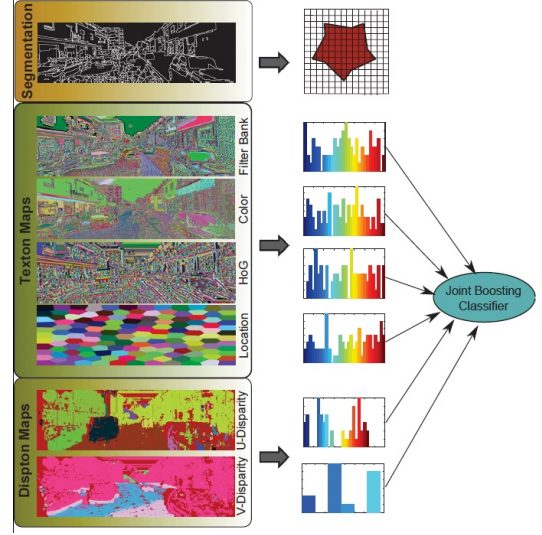


Fig. 2. The Road Recognition process. For each  $sp_i$  is constructed the multi-normalized histogram from the Texton maps and the Dispton maps, which are classified with Joint Boosting algorithm, explained in Section III-C

*classifiers*, simultaneously selecting discriminative features. Different from *TextonBoost* that works with the shape filters modeled in the weak classifiers, this approach builds a multi-normalized-histogram of the superpixel features. Representative *Textons* and *Disptons* histograms for all segments are generated where each bin represents the cumulative number of a given cluster in their respective regions, as can be seen in Figure 2. Thereby, the Joint Boosting algorithm is an additive model of the form  $H(c_i) = \sum_{m=1}^M h_m(c_i)$ , that sum the classification confidence of  $M$  joint weak classifiers. In this case,  $H(c_i)$  represents the strong learned classifier obtained by the  $\chi^2$  metric comparisons of the weak classifiers' histogram with the response's histogram obtained from the image. The output of these comparisons are modeled as a decision stump as shown in Equation(1):

$$h(c_i) = \begin{cases} a\delta(d(rh, sp_i) > \theta) + b & , \text{ if } c_i = 1 \\ \kappa_{c_i} & , \text{ otherwise} \end{cases} \quad (1)$$

In equation (1),  $rh$  represents the random histogram comparator and  $sp_i$  represents the multi-normalized histogram of the superpixel  $i$ . The  $d(\cdot)$  is the  $\chi^2$  function and  $\delta(\cdot)$  is a 0-1 indicator function. The weak learner gives  $h(c_i) \in \{a+b, b\}$  depending on the comparison of  $d(rh, sp)$  with a threshold  $\theta$ . The constant  $\kappa_{c_i}$  ensures asymmetrical values of positive and negative training examples. The result is the probability given by (2):

$$P = \frac{1}{Z} \exp^{-H(c_i)} \quad (2)$$

where  $Z$  is a normalization factor.

#### IV. EXPERIMENTAL RESULTS

In this section we present the results of our experiments, using the Urban Kitti-Road<sup>5</sup> dataset. The validation platform were implemented in C++, and the experiments were executed in one equipment with an Intel Xeon E5-1650 processor with 3.20Ghz and with 16Gb DDR3, running the version 7 of the Windows OS.

In brief, this dataset consists of  $\simeq 600$  frames (375x1242 px) recording from five different days and containing relatively low traffic density [12], representing a typical road scene in inner-city. Data is categorized in three sets having each one a subset of training images and a subset of test images exclusively for evaluation performance. The first one is urban unmarked (UU) with 98 images for training and 100 images for test, followed by urban marked two-way road (UM) with 95 and 96 images, and urban marked multi-lane road (UMM) with 96 and 94 images. The experiments in this work were based in the perspective and metric space for all categories, which each one had its own learning process. Thus, a sample set was built for each one, extracting  $\simeq 12.8E + 4$  samples from the UU image training set,  $\simeq 12.6E + 4$  samples from the UM image training set and  $\simeq 11.7E + 4$  samples from the UMM image training set. These sets were used in both algorithms to perform the feature's descriptor using the ANN approach and the multi-normalized histogram for the Joint Boosting approach, all of them using the superpixel area.

The methodology to train the ANN algorithm uses the cross validation method, splitting the training images in training, validation and test subsets, containing  $\simeq 15\%$ ,  $\simeq 7\%$  and  $\simeq 78\%$  respectively. The time to train each category takes around 24 hours and the classification time for a single image takes around 3 seconds. It should be mentioned that some adaptations from the first version proposed in [11] were required: (i) it does not use the moving average technique, because it is applicable only for image sequences, (ii) the learning process does not use the strategy of training with subclass (shadow area, normal area and land marks), where it would improve the final result of road detection. In the case of the learning process to Joint Boosting algorithm, the training images were splitted into training subset containing  $\simeq 40\%$ , and the test subset with  $\simeq 60\%$ . The time to train each category of this approach takes around 72 hours and the classification time for a single image takes around 2.5 minutes.

For all quantitative evaluation, the baseline to road area is provided as a lower bound [12], by averaging all ground truth road maps from the present validation set. The Table I presents the quantitative evaluation for the UU category. As can be seen, the HistonBoost approach reached a improvement of 24.28% from ANN and 4.13% from Baseline in the perspective space. Based on the metric space using the testing set, the same approach takes 73.51%, reaching a gain of 19.34% and 4,02% from ANN and Baseline. The



Fig. 3. The result process for Urban Unmarked (UU) Perspective image. The first row shows the original image into training set. The second row presents the ground truth, followed by the next two rows that are the HistonBoost result and Artificial Neural Network result.

qualitative result for this challenging category can be seen in the Figure 3, where it presents the classification process on perspective space using the training set.

TABLE I  
RESULTS [%] OF PIXEL-BASED FOR URBAN UNMARKED ROAD AREA EVALUATION.

Perspective space - Training set						
	$F_{max}$	AP	Prec.	Recall	FPR	FNR
BaseLine	80.79	<b>86.13</b>	79.00	82.67	3.42	17.33
ANN	60.64	41.67	44.49	<b>95.22</b>	18.51	<b>4.78</b>
HistonBoost	<b>84.92</b>	71.36	<b>84.22</b>	85.63	<b>2.50</b>	14.37
Metric space - Testing set						
	$F_{max}$	AP	Prec.	Recall	FPR	FNR
BaseLine	69.49	<b>73.84</b>	65.73	73.70	12.78	26.30
ANN	54.17	36.86	39.50	<b>86.19</b>	43.92	<b>13.81</b>
HistonBoost	<b>73.51</b>	63.07	<b>77.36</b>	70.03	<b>6.82</b>	29.97

With respect to UM category, Table II shows that HistonBoost reached a improvement of 19.04% from ANN and from Baseline this percentage is approximately 1.00% in the perspective space. In case of the metric space using the testing set, this approach reached 83.71%, reaching a gain of 21.07% and 1,18% from ANN and Baseline. The qualitative result for this category can be seen in the Figure 4.

TABLE II  
RESULTS [%] OF PIXEL-BASED FOR URBAN MARKED ROAD AREA EVALUATION.

Perspective space - Training set						
	$F_{max}$	AP	Prec.	Recall	FPR	FNR
BaseLine	89.72	<b>93.15</b>	89.19	90.26	2.11	9.74
ANN	71.54	53.31	57.02	<b>95.97</b>	13.94	<b>4.03</b>
HistonBoost	<b>90.58</b>	83.47	<b>90.20</b>	90.96	<b>1.90</b>	9.04
Metric space - Testing set						
	$F_{max}$	AP	Prec.	Recall	FPR	FNR
BaseLine	82.53	<b>85.59</b>	79.24	<b>86.11</b>	10.41	<b>13.89</b>
ANN	62.64	46.80	50.18	83.34	38.21	16.66
HistonBoost	<b>83.71</b>	73.31	<b>82.58</b>	84.87	<b>8.27</b>	15.13

Using the UMM category, which is the less complex if compared to others, we can highlight the ANN approach due to its response in the testing set overcoming the baseline approach with 4.92%. The result for this category can be seen in Figure 5. Compared to other approaches, ANN outcomes presents the worst results. It can be explained by the higher

<sup>5</sup>[http://www.cvlibs.net/datasets/kitti/eval\\_road.php](http://www.cvlibs.net/datasets/kitti/eval_road.php)



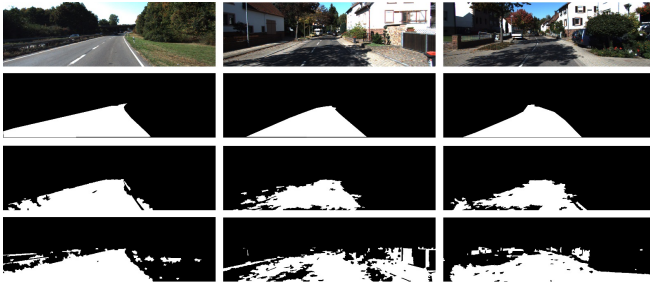


Fig. 4. The result process for Urban Marked (UM) Perspective image. The first row shows the original image into training set. The second row presents the ground truth, followed by the next two rows that are the HistonBoost result and Artificial Neural Network result.

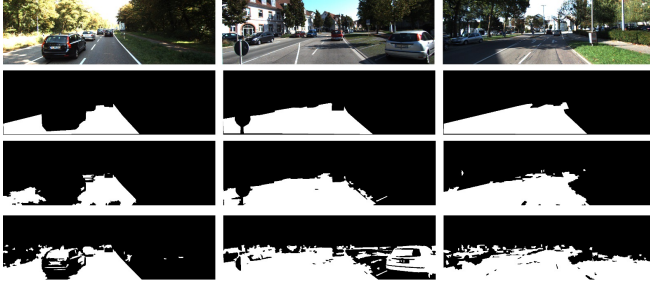


Fig. 5. The result process for Urban Marked Multi-lane (UMM) Perspective image. The first row shows the original image into training set. The second row presents the ground truth, followed by the next two rows that are the HistonBoost result and Artificial Neural Network result.

complexity of the Kitti-road dataset, or, the training process used for ANN was not adequate, having low expressiveness if observed the strategy of subclasses used in [11].

TABLE III

RESULTS [%] OF PIXEL-BASED FOR URBAN MARKED MULTI-LANE ROAD AREA EVALUATION.

Perspective space - Training set						
	$F_{max}$	AP	Prec.	Recall	FPR	FNR
BaseLine	82.81	<b>89.21</b>	77.54	88.86	8.02	11.14
ANN	82.04	68.79	73.30	93.14	10.58	6.86
HistonBoost	<b>88.72</b>	78.45	<b>83.92</b>	<b>94.11</b>	<b>5.62</b>	<b>5.89</b>
Metric space - Testing set						
	$F_{max}$	AP	Prec.	Recall	FPR	FNR
BaseLine	76.17	78.42	65.02	91.95	57.89	8.05
ANN	81.09	68.93	70.43	<b>95.56</b>	46.94	<b>4.44</b>
HistonBoost	<b>87.70</b>	<b>81.59</b>	<b>84.36</b>	91.32	<b>19.81</b>	8.68

To conclude the evaluation process, Table IV presents the final results merging all categories. The HistonBoost approach presents 83.41% of this challenging urban Kitti-road benchmark, where can be considerable a satisfactory result given the complexity of the scenarios. Figure 6 shows the qualitative results using the testing set and demonstrating the performance in the metric space, denoted by Bird's Eye View (BEV). The road detection results for HistonBoost and ANN approaches are demonstrated in red.

## V. CONCLUSIONS AND FUTURE WORKS

This proposed work presents the meaningful performance analysis using two distinct road detect algorithms in inner-

TABLE IV

RESULTS [%] OF PIXEL-BASED FOR COMPLETE URBAN ROAD AREA EVALUATION, PERFORMED ON THE METRIC SPACE (BEV).

	$F_{max}$	AP	Prec.	Recall	FPR	FNR
BaseLine	75.61	<b>79.72</b>	68.93	83.73	21.73	16.27
ANN	68.12	51.52	54.85	<b>89.85</b>	42.59	<b>10.15</b>
HistonBoost	<b>83.41</b>	74.06	<b>82.39</b>	84.46	<b>10.39</b>	15.54

city, showing the pros and cons of each approach. The first approach constructs a feature's descriptor extracting statistical measures of 2D and 3D information to be classified using an Artificial Neural Network. The second approach construct a multi-normalized histogram feature's descriptor to be recognized by a Joint Boosting classifier.

Based on the experimental results, the performance analysis demonstrates that the HistonBoost approach copes better than the ANN approach for all dataset in the perspective and metric space. The main advantages of this method are: (i) The union of Textons and Disptons to compose the feature's descriptor shows promising applicability; (i) The Joint Boosting algorithm is able to build a satisfactory discriminative model to road pattern in complex scenarios; (iii) The Module independency of this approach leverages the usage of parallelization and improvement on resources usage. However, the processing time is still an issue, if compared with other approaches.

In general, The high level of complexity to build and test a recognition model is due to the different characteristics present in inner-city environments. We believe that an alternative to model a complex road pattern should take into account the learning of intra-classes pattern derived from the main class such as normal area, shadow area, horizontal land mark if exist, area with influence of sun, etc. Thereby, an strategy using any kind of unsupervised classifier to leave arise clusters of these intra-classes automatically, and after that, merging this result process in a supervised method could be interesting to improve the result of road recognition.

Taking in accounts these considerations and analysis, we are still working on improvements to reduce the processing time using the GPU architecture. We are also working on Self-Organizing Maps [23], in order to better discriminate the road pattern and extending the recognition for different classes such as vehicles, builds, sidewalks, etc, to improve for example the classification mistakes between road and non-road.

## REFERENCES

- [1] P. Karasev, M. Serrano, P. Vela, and A. Tannenbaum, "Depth invariant visual servoing," in *Decision and Control and European Control Conference (CDC-ECC)*, 2011, pp. 4992–4998.
- [2] N. Rawashdeh and H. Jasim, "Multi-sensor input path planning for an autonomous ground vehicle," in *Mechatronics and its Applications (ISMA)*, 2013, pp. 1–6.
- [3] T. Khnl, F. Kummert, and J. Fritsch, "Monocular road segmentation using slow feature analysis," in *Intelligent Vehicles Symposium*. IEEE, 2011, pp. 800–806.
- [4] C. Tan, T. Hong, T. Chang, and M. Shneier, "Color model-based real-time learning for road following," in *Intelligent Transportation Systems Conference, ITSC*, 2006, pp. 939–944.

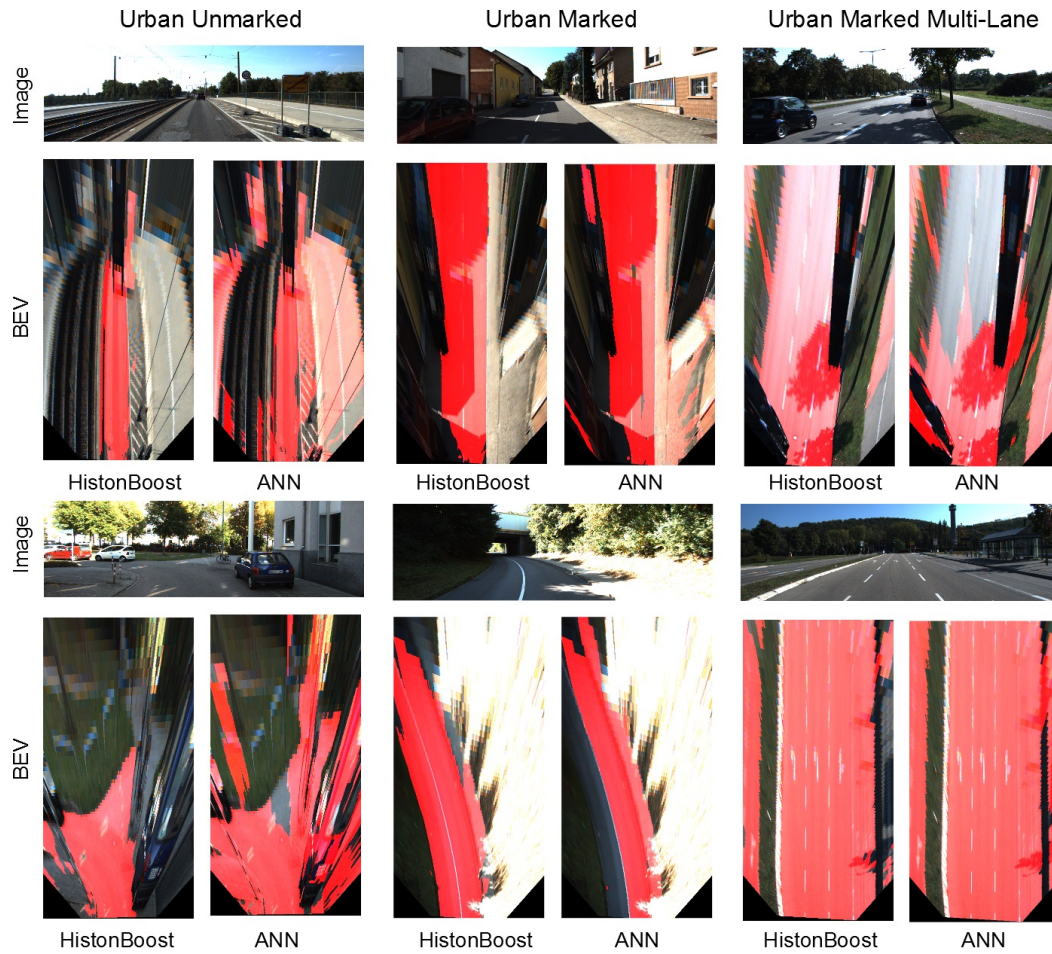


Fig. 6. The Classification results using the testing set. The road detection for each approach is highlighted in red.

- [5] J. Alvarez, T. Gevers, and A. Lopez, "Learning photometric invariance from diversified color model ensembles," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 565–572.
- [6] J. Alvarez and A. Lopez, "Road detection based on illuminant invariance," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 1, pp. 184–193, 2011.
- [7] 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, 2004, pp. I-470–I-477 Vol.1.
- [8] H. Kong, J.-Y. Audibert, and J. Ponce, "General road detection from a single image," *Image Processing, IEEE Transactions on*, vol. 19, no. 8, pp. 2211–2220, 2010.
- [9] N. Soquet, D. Aubert, and N. Hautiere, "Road segmentation supervised by an extended v-disparity algorithm for autonomous navigation," in *Proceedings of the IEEE Symposium on Intelligent Vehicles*, 2007, pp. 160–165.
- [10] P. Sturgess, K. Alahari, L. Ladicky, and P. H. S. Torr, "Combining appearance and structure from motion features for road scene understanding," in *BMVC. British Machine Vision Association*, 2009.
- [11] 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.
- [12] 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.
- [13] E. R. Dougherty and R. A. Lotufo, *Hands-on Morphological Image Processing (SPIE Tutorial Texts in Optical Engineering Vol. TT59)*. SPIE Publications, July 2003.
- [14] D. A. Lima, G. B. Vitor, A. C. Victorino, and J. V. Ferreira, "A disparity map refinement to enhance weakly-textured urban environment data," in *International Conference on Advanced Robotics (ICAR), 2013 IEEE*, 2013.
- [15] O. Faugeras, *Three-dimensional computer vision: A geometric view point*. Cambridge: MIT Press, 1993.
- [16] R. Labayrade, D. Aubert, and J. P. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation," in *Proceedings of the IEEE Symposium on Intelligent Vehicles*, vol. 2, 2002, pp. 646–651.
- [17] P. Y. Shinzato and D. F. Wolf, "A road following approach using artificial neural networks combinations," *J. Intell. Robotics Syst.*, vol. 62, no. 3-4, pp. 527–546, June 2011. [Online]. Available: <http://dx.doi.org/10.1007/s10846-010-9463-2>
- [18] L. Ladicky, C. Russell, P. Kohli, and P. H. S. Torr, "Associative hierarchical crfs for object class image segmentation," in *Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 739–746.
- [19] J. Shotton, J. M. Winn, C. Rother, and A. Criminisi, "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," *International Journal of Computer Vision*, vol. 81, no. 1, pp. 2–23, 2009.
- [20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *In CVPR*, 2005, pp. 886–893.
- [21] A. Torralba, K. Murphy, and W. Freeman, "Sharing features: efficient boosting procedures for multiclass object detection," in *Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2004, pp. II-762–II-769 Vol.2.
- [22] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," 2007.
- [23] T. Kohonen, *Self-organizing maps*, 3rd ed., ser. Springer series in information sciences, 30. Berlin: Springer, December 2001.