A Comparative Analysis of Decision Trees Based Classifiers for Road Detection in Urban Environments

C. Fernández, R. Izquierdo, D. F. Llorca, M. A. Sotelo

***Abstract*— In this paper a comparative analysis of decision trees based classifiers is presented. Two different approaches are presented, the first one is a speficic classifier depending on the type of scene. The second one is a general classifier for every type of scene. Both approaches are trained with a set of features that enclose texture, color, shadows, vegetation and other 2D features. As well as 2D features, 3D features are taken into account, such as normals, curvatures and heights with respect to the ground plane. Several tests are made on five different classifiers to get the best parameters configuration and obtain the importance of each features in the final classification. In order to compare the results of this paper with the state of the art, the system has been tested on the KITTI Benchmark public dataset.**

I. INTRODUCTION AND RELATED WORK

Autonomous driving is a high priority issue on the research of car makers and research centers. In recent years, Advanced Driving Assistance Systems (ADAS) have been deployed in mass-produced cars. ADAS are based in different technologies, such as RADAR, LIDAR and vision. As an example, in highways applications RADAR is consolidated for Automatic Cruise Control (ACC) systems, however urban scenarios require a precise detection of the scene and vision is consolidated because of the low cost and rich information provided. Detection of free space is very important for other tasks in autonomous navigation, such as path planning. Urban scenarios are particularly challenging because of the large variety of street configurations and evironment conditions. For example, drivable space sometimes is only limited by small curbs, visibility of road edges can be occluded and illumination conditions suddenly change.

A review of related literature reveals that color and texture are potential features to characterize the road [1], [2]. Challenging situations are frequently caused by shadows when the scene has both shadowed and nonshadowed areas. In [3], an illuminant invariant feature is combined with a model-based classifier to obtain a system robust to shadows. The methods described in [4], [5] and [6] reveals that spatial information enhances local classification decisions and therefore road detection. Monolothic classifiers such as Artificial Neural Networks (ANN) and Supported Vector Machine (SVM) have been utilized in several applications for data classification [7], [8] and [9], however recent approaches based on weak classifiers outperform traditional monolithic classifiers for the road detection problem. In [6] the authors propose to train a GentleBoost with the selected features. In other cases, multi-normalized histogram from a set of features are used to train a joint boosting classifier [10].

The rest of the paper is structured as follows: section II presents a general description of the system. The features set and the classifiers used for the classification stage are described in section III and IV respectively. Results and discussion are presented in section V. Finally, we analyze our conclusions and future work in section VI.

C. Fernández, R. Izquierdo, D. F. Llorca, and M. A. Sotelo are with the Computer Engineering Department, Polytechnic School, University of Alcalá, Madrid, Spain. email: carlos.fernandez, llorca, sotelo@aut.uah.es.

II. SYSTEM DESCRIPTION

Since we focus on urban environments, the performance of the system is evaluated using the public dataset: KITTI Vision Benchmark Suite [11]. The dataset provides images and information of urban scenarios from different types of sensors, such as monochrome and color cameras, multilayer LIDAR, GPS and IMU. For this paper, only information from cameras are processed. The height and base line of the stereo cameras are 1.65 m and 0.54 m respectively and the cameras have 1.4 Mpx (Point Grey Flea 2).

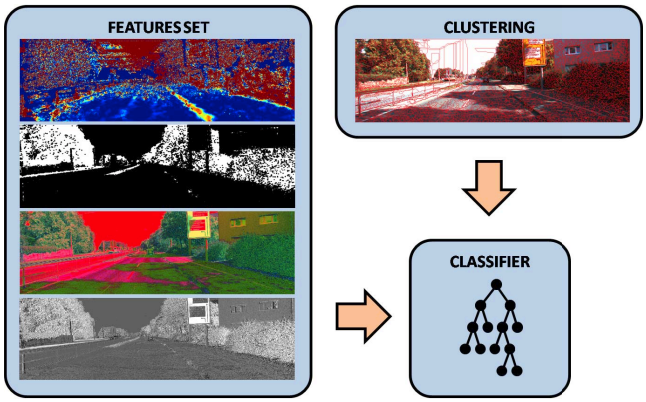

Fig. 1. System diagram. 3D features are extracted from an environment reconstruction using stereo cameras. The clustering stage decreases the number of samples and therefore increases the speed of the system. Finally, the feature set is used in a decision tree based classifier.

As mentioned in section I, precise understanding of urban scenarios is crucial for autonomous driving. In the proposed method, we classify the scene in two categories: non road and road. In order to get a better performance, different classifiers are compared: Boosting Discrete (BoostD), Boosting Gentle (BoostG), Extremely Randomized Trees (ERT), Random Trees (RT) and Decision Trees (DT). All of them are trained with several parameters and only the best combination of them are compared. The feature vector is composed of 2D

978-1-4673-6596-3/15 $31.00 © 2015 IEEE
DOI 10.1109/ITSC.2015.122

and 3D information. 3D information is very discriminative to detect the ground plane and big obstacles such as vehicles and buildings, however, this information is not always well estimated and the system also should take into account 2D features such as color, texture and road markings.

III. FEATURES DESCRIPTION

A. 3D Features

The 3D features are extracted from a point cloud obtained using the Semi Global Matching (SGM) algorithm [12]. These features are XYZ coordinates, normals, height with respect to the ground plane and curvatures. The normal vector is estimated with a plane tangent to the surface using the least-square plane fitting method. The solution is reduced to an analysis of the eigenvectors and eigenvalues of a covariance matrix C created from the nearest neighbors of the query point p_i.

$$C = \frac{1}{k}\sum_{i=1}^{k}(p_i - \overline{p})\cdot(p_i - \overline{p})^T \tag{1}$$

Normals provide information of the orientation of the surface. However, there is another feature more robust and stable than normals: surface curvature. The surface curvature estimation method was presented in [13] and it has been also used in [14] for free space detection. The curvature describes the variation along the surface normal and it varies between 0 and 1, where low values correspond to flat surfaces. For each point p, the nearest neighbors (NN) p_i in a surrounding area defined by a radius R are selected. These points are used to create a weighted covariance matrix, where k denotes the number of NN.

$$\bar{p} = \frac{1}{k}\sum_{i=1}^{k} p_i \;;\; \mu = \frac{1}{k}\sum_{i=1}^{k}|p - p_i| \tag{2}$$

$$w_i = \begin{cases} exp\left(-\frac{(p-p_i)^2}{\mu^2}\right) & if\,|p-p_i| \geq \mu \\ 1 & otherwise \end{cases} \tag{3}$$

$$C = \sum_{i=1}^{k} w_i \cdot (p_i - \bar{p})\cdot(p_i - \bar{p})^T \tag{4}$$

This method has the advantage of taking into account the distance from every outlier to the query point using the robust estimator above. The eigenvector V and eigenvalues λ of C are computed as $C \cdot V = \lambda \cdot V$. The curvature measure γ_z^p is defined by equation 5, where $\lambda_x \leq \lambda_y \leq \lambda_z$ are the eigenvalues of the covariance matrix C.

In Figure 2, curvature values are represented in a color scale. Finally, the ground plane is estimated using RANSAC in the point cloud and the height of every point with respect to the plane is computed.

$$\gamma_z^p = \frac{\lambda_z}{\lambda_x + \lambda_y + \lambda_z} \tag{5}$$

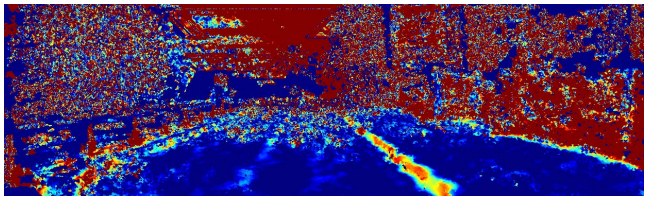

Fig. 2. Representation of curvature values in a color scale.

B. 2D Features

The 2D features set is composed of a HSV color image, a vegetation segmentation method, a road marking detection function, an illuminant invariant image, a shadow detection function and a texture anisotropy image, see Figure 3

1) Color: Instead of using only the HSV channels as a feature, a more elaborated feature is created for vegetation segmentation. An area of the hue channel is selected to segment the green areas of the scene and some filtering is applied to the resulted image.

2) Road marking: Furthermore, the road marking usually provides relevant information about the road limits, specially in urban environments. The proposed road marking detection method is based on state of the art techniques. However, we provide a brief description for completeness purpose. As explained in [15], a median filter is applied to the input image. The window size of the median filter needs to be twice larger than the road marking. If the road marking is larger than the window, for example in a zebra crossing, the border is well detected but the areas inside the zebra crossing are not. In order to keep the window size constant, a bird-eye view of the scene is reconstructed. An adaptable threshold is then applied to the input image. After that, both images are subtracted and the final result is filtered to remove some noise.

3) Illuminant invariant: Urban scenarios are strongly affected by shadows. Road detection is a challenge specially when the shadows are from trees because their shadow has an irregular shape with holes inside. The illuminant invariant image provides information of the environment not affected by shadows [16]. In the process to obtain the illuminant invariant image, chromaticity and illumination are splited from the original image. The illumination of the scene and the grey image are the inputs of a specific function to detect the shadows of the scene. When the shadows are segmented, a special image processing algorithm can be applied to this areas because the lighted surface and the same surface affected by shadows look very different.

4) Texture anisotropy: The strength of texture anisotropy becomes a powerful feature to distinguish the homogeneous areas of the image. As explained in [17], firstly, the first derivatives f_x f_y, the local mean μ_x and the covariance σ_{xx} are computed using equations 6 and 7, where Σ' is a normalized weighting Gaussian function. Following the same procedure, μ_y, σ_{xy}, σ_{yx} and σ_{yy} are computed to create a gradient covariance matrix M, see equation 8.

$$\mu_x = \sum_{i=-n}^{n} \sum_{j=-n}^{n} w(i,j) f_x(x+i, y+j) \tag{6}$$

$$\sigma_{xx} = \sum_{i=-n}^{n} \sum_{j=-n}^{n} w(i,j) f_x^2(x+i, y+j) - \mu_x^2 \tag{7}$$

$$\Sigma' = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} \tag{8}$$

The eigenvalues λ_1 and λ_2 of Σ' are both small in an homogeneous region, however, in the proximity of edges both eigenvalues are large. To characterize the relationship between both eigenvalues, we define the strength of texture anisotropy s following equation 9, where $\lambda_1 > \lambda_2$ and c is the normalization factor. The resulted value is a real value between 0 and 1 where homogeneous regions have small values.

$$s = \frac{\lambda_1}{c} \left(\frac{\sqrt{\lambda_1^2 - \lambda_2^2}}{\lambda_1} \right)^2 \tag{9}$$

5) Other features: In addition to the features described before, the pixel position in image coordinates is appended to the feature vector.

IV. CLASSIFICATION

The classification process is performed over superpixels instead of every individual pixel. The use of superpixels reduces 46 times the number of samples and consequently the training and prediction complexity. In addition, all pixels in a superpixel are most likely uniform, therefore noise is mitigated and the structure of the objects is preserved. The clustering method is based on Watershed Transform. This function requires to roughly outline the desired regions. These markers are seeds of the future image regions. In order to obtain some flexibility in the segmentation level, different seed images are used. The first seed image is the texture anisotropy computed in section III-B.4 and the following images are the same image after some morphological filters. Finally, the resulted images are added to obtain the clusters depicted in Figure 4. Every cluster is a sample for the training stage. Because of the homogeneity of the superpixels, the mean value of the pixels is computed for the complete features set.

As explained in section I, boosting techniques are becoming very relevant in the road classification problem. This technique combines the performance of many weak classifiers to produce a strong classifier. The weak classifier is computationally fast and it is usually a decision tree. Instead of using decision trees as weak classifiers, they also can be used for classification, where each tree leaf is marked with a class label and multible leaves may have the same label. Random trees is a collection of decision trees, because of that, is also known as random forest. Every decision tree takes the input feature vector, classifies it and the forest output is the class label that received more votes. During the training stage, at each tree node, a random subset of features

(a) Input image

(b) HSV

(c) Vegetation

(d) Road marking

(e) Illuminant invariant image

(f) Shadow detection result

(g) Texture anisotropy

Fig. 3. Graphic representation of the 2D features

(a) Left: Original image. Right: Seed image

(b) Clustering result

Fig. 4. Superpixels used to classify the scene

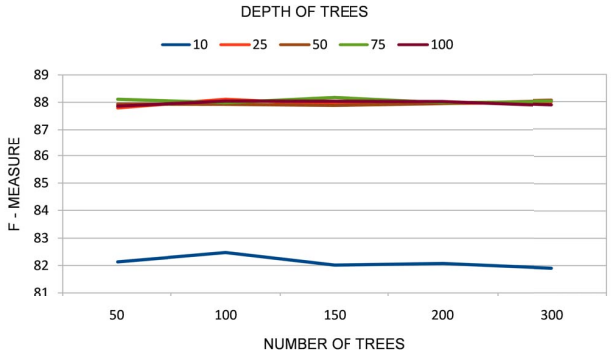

Fig. 5. F-measure of a Random Tree (RT) classifier varying depth and number of trees during the training stage.

are used to find the best split value, in contrast, extremely randomized trees choose the feature index and the split value randomly.

In order to find the best classifier for the road detection problem in urban scenarios, the following classifiers are compared: Boosting Discrete (BoostD), Boosting Gentle (BoostG), Extremely Randomized Trees (ERT), Random Trees (RT) and Decision Trees (DT).

V. RESULTS

The metrics for the evaluation of the classifiers are: Precision, recall, accuracy and F-measure, where $\beta = 1$ for an harmonic mean of F-measure.

$$F-Measure = (1+\beta^2)\frac{Precision \cdot Recall}{\beta^2 Precision + Recall} \qquad (10)$$

In our experiments the classifiers have 2 labels: non road and road. In the training stage, the scenes are divided in 3 groups: urban marked (UM), urban multiple marked lanes (UMM) and urban unmarked (UU). Our first approach is to train 3 classifiers independently for every scene and modify the training parameters to get the influence of them on the F-measure. For every scene, $2/3$ of the samples are used for training and $1/3$ for testing. The most important parameters to adjust are the number of trees, the maximum depth of each tree and the cost of a missclassification for a specific label. In Figure 5, a comparison of the F-measure (in perspective space) against the number of trees and the depth of each tree is shown. In this case, it is significant how the classifier improves 6% with depths greater than 10, nevertheless greater depths have no effect and the number of trees is not relevant.

Our second approach is to train only one classifier for all the scenes with the same samples used before. The statistics explained before are computed in a perspective point of view. This metrics evaluate the quality of the classification in pixels, however in a real scenario of autonomous vehicle, it is more important the behavior in a metric space of the closer 46 meters in front of the vehicle [18]. For example, if the sky is wrongly classified it is not relevant in the real application because only the closer meters are taken into account. Comparing the results of the specific classifiers versus a general one, the average performance is very similar in perspective evaluation. Comparing perspective and metric evaluation for the best cases, the perspective evaluation achieves a F-measure of 5% higher than in metric. In perspective, every pixel has the same weight in the final result, on the contrary, in metric evaluation the further pixels become more relevant for the final result. 3D information from stereo is only reliable up to 30 meters. The reason of a lower performance in metric evaluation is missclassifications on further distance caused by unreliable 3D information.

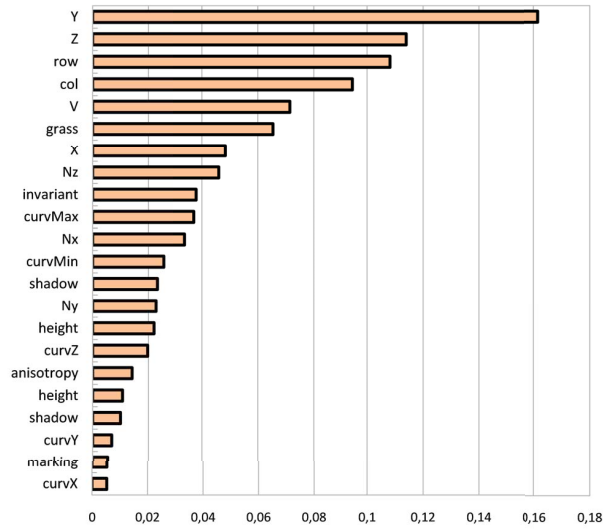

Fig. 6. Weight of the features in the final classification response.

The weights assigned for each feature during the training stage reveal that 3D features (Y and Z coordinates) and its 2D representation (column and row) are the most discriminant features. However some of the other 2D features are still important such as the grey value of (HSV) or the vegetation, see Figure 6. The representation of the classification results in the image, see Figure 7, yields that some of the missclas-

TABLE I

PERFORMANCE OF CLASSIFIERS.

URBAN MARKED										
	Perspective			Metric						
Method	F-measure	Precision	Recall	F-measure	Precision	Recall	Depth	Trees	w_{NR}	w_R
DT	79.32	86.51	73.24	76.72	91.58	66.00	1	-	10	1
RT	81.49	92.04	73.11	78.13	91.39	68.24	75	200	10	1
ERT	81.44	91.93	73.10	78.04	91.22	68.20	100	50	10	1
BoostG	85.48	78.88	93.30	80.29	70.48	93.28	1	750	1	1
BoostD	85.24	82.01	88.73	85.15	79.69	91.41	5	750	1	10
URBAN MULTIPLE MARKED LANES										
	Perspective			Metric						
Method	F-measure	Precision	Recall	F-measure	Precision	Recall	Depth	Trees	w_{NR}	w_R
DT	80.67	69.45	96.21	83.22	78.50	88.56	1	-	1	1
RT	90.53	94.65	86.76	86.91	90.86	83.28	75	150	10	1
ERT	90.40	94.51	86.63	86.78	90.72	83.16	25	200	10	1
BoostG	93.30	93.55	93.04	89.46	94.82	84.67	25	250	1	10
BoostD	93.09	92.96	93.23	89.37	89.24	89.50	25	750	1	10
URBAN UNMARKED										
	Perspective			Metric						
Method	F-measure	Precision	Recall	F-measure	Precision	Recall	Depth	Trees	w_{NR}	w_R
DT	71.04	75.11	67.39	56.27	73.62	45.54	1	-	10	1
RT	79.04	88.09	71.68	41.74	88.59	27.30	100	100	10	1
ERT	79.13	89.37	71.00	41.64	86.07	27.17	50	100	10	1
BoostG	81.72	84.71	78.92	57.68	71.81	48.20	750	5	1	1
BoostD	82.08	81.00	83.18	62.60	56.33	70.43	250	5	1	1
ALL SCENES										
	Perspective			Metric						
Method	F-measure	Precision	Recall	F-measure	Precision	Recall	Depth	Trees	w_{NR}	w_R
DT	74.17	83.35	66.81	67.02	71.98	70.47	1	-	10	1
RT	82.27	92.11	74.33	56.76	57.91	77.96	150	50	10	1
ERT	82.29	92.05	74.41	56.76	57.91	77.96	300	25	10	1
BoostG	86.58	86.14	87.03	79.29	67.83	90.39	250	25	1	10
BoostD	87.06	86.75	87.38	79.05	66.98	92.70	1500	25	1	1

sifications are sidewalks due to small curbs. Furthermore, in challenging urban scenarios the limit of a drivable area is difficult to distinguish from the non drivable area, such as a cyclist lane. In some cases the limit is just a road marking and the texture and the 3D features looks very similar. It is remarkable that road markings have a low weight in the final response. For this reason we think that this feature can be very useful in a higher level stage for scene understanding.

VI. CONCLUSIONS AND FUTURE WORK

A road classification method is presented in this paper using a wide range of 3D and 2D features. The comparison of the trained classifiers reveals that ERT and RT are very similar to each other but slightly worse than boosting. On the one hand, GentleBoost is the training technique that best generalizes the road structure in complex urban scenearios. On the other hand, Decision Trees are 10% worse than boosting in the road detection problem. In order to solve the most challenging urban scenarios, the system requires a higher level module that uses the results presented in this paper as inputs. We think that road marking provides very important information when they are available, therefore instead of using this feature for the road classification, we will include it in a higher level module. In addition, it is planned to include the curb detection method presented in [19] to separate the road from the sidewalk. For future work, traffic rules will be also taken into account to establish the road limits.

VII. ACKNOWLEDGMENTS

This work was supported by the Research Grants DIS-ADAPT SPIP2014-1300 (General Traffic Division of Spain), IMPROVE DPI2014-59276-R (Spanish Ministry of Economy), and SEGVAUTO-TRIES-CM S2013/MIT-2713 (Community of Madrid).

REFERENCES

[1] M. A. Sotelo, F. J. Rodriguez, L. Magdalena, L. M. Bergasa, and L. Boquete, "A color vision-based lane tracking system for autonomous driving on unmarked roads," *Autonomous Robots*, vol. 16, no. 1, pp. 95–116, 2004.

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

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

[4] J. Alvarez, T. Gevers, and A. Lopez, "3d scene priors for road detection," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, June 2010, pp. 57–64.

[5] A. Hata, F. Osorio, and D. Wolf, "Robust curb detection and vehicle localization in urban environments," in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, June 2014, pp. 1257–1262.

[6] J. Fritsch, T. Kuhnl, and F. Kummert, "Monocular road terrain detection by combining visual and spatial information," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 4, pp. 1586–1596, Aug 2014.

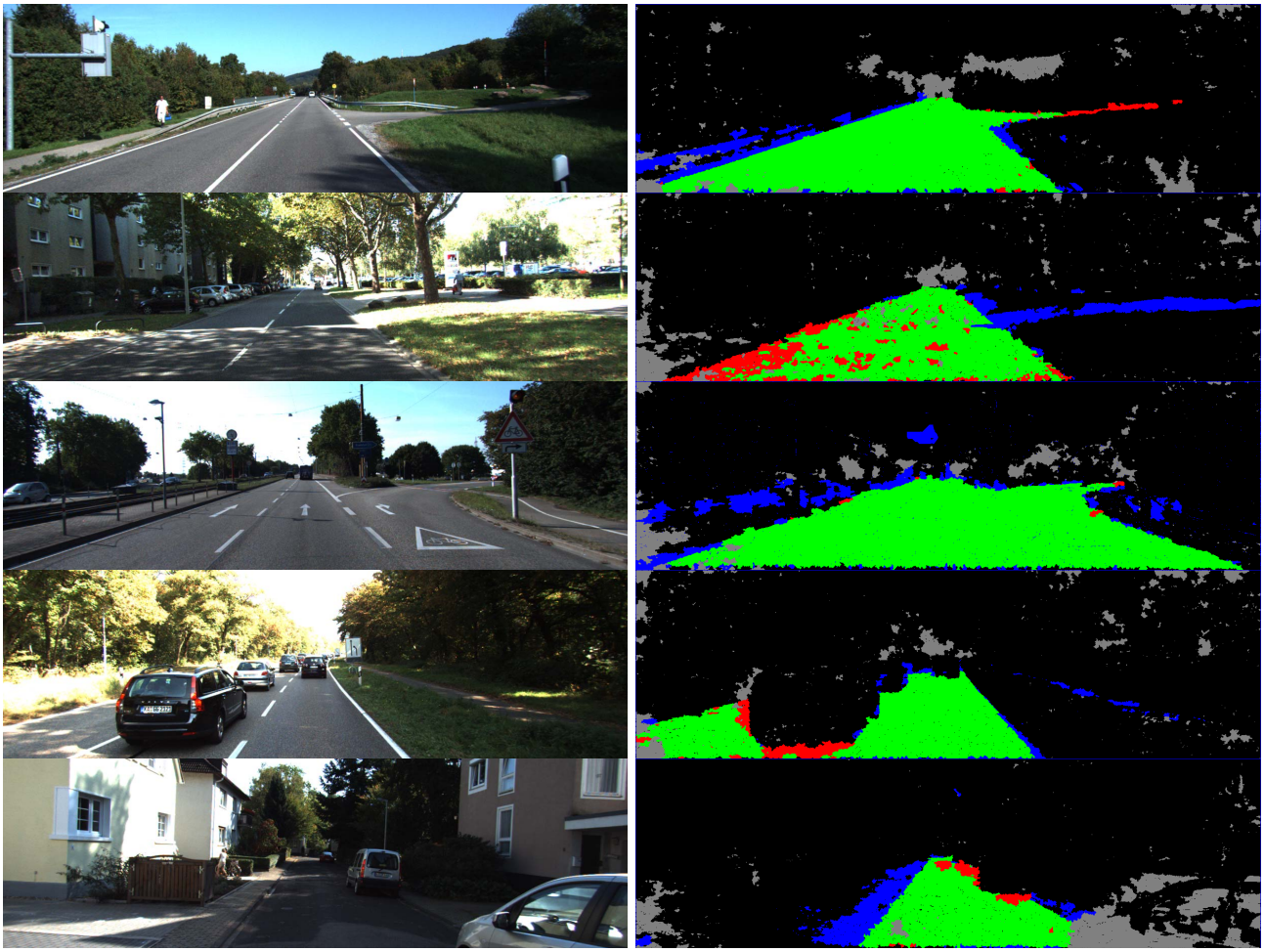

Fig. 7. Final results in different urban scenarios. Scenes affected by shadows, residential areas and roads with traffic are analyzed. TP are painted in green, FP in blue, FN in red and the grey areas correspond to unknown 3D information.

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

[8] J. Alvarez, Y. LeCun, T. Gevers, and A. Lopez, "Semantic road segmentation via multi-scale ensembles of learned features," vol. 7584, pp. 586–595, 2012.

[9] I. Alonso, D. Llorca, M. Sotelo, L. Bergasa, P. Revenga de Toro, J. Nuevo, M. Ocana, and M. Garrido, "Combination of feature extraction methods for svm pedestrian detection," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8, no. 2, pp. 292–307, June 2007.

[10] G. Vitor, A. Victorino, and J. Ferreira, "Comprehensive performance analysis of road detection algorithms using the common urban kitti-road benchmark," in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, June 2014, pp. 19–24.

[11] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *International Journal of Robotics Research (IJRR)*, 2013.

[12] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 807–814.

[13] M. Pauly, M. Gross, and L. Kobbelt, "Efficient simplification of point-sampled surfaces," in *Visualization, 2002. VIS 2002. IEEE*, Nov 2002, pp. 163–170.

[14] C. Fernandez, R. Izquierdo, D. Llorca, and M. Sotelo, "Road curb and lanes detection for autonomous driving on urban scenarios," in *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, Oct 2014, pp. 1964–1969.

[15] P. Foucher, Y. Sebsadji, J. P. Tarel, P. Charbonnier, and P. Nicolle, "Detection and recognition of urban road markings using images," in *14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Oct 2011, pp. 1747–1752.

[16] G. Finlayson, S. Hordley, C. Lu, and M. Drew, "On the removal of shadows from images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 59–68, Jan 2006.

[17] J. Zhang and H.-H. Nagel, "Texture-based segmentation of road images," in *Intelligent Vehicles '94 Symposium, Proceedings of the*, Oct 1994, pp. 260–265.

[18] 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.

[19] C. Fernandez, R. Izquierdo, D. Llorca, and M. Sotelo, "Curvature-based curb detection method in urban environments using stereo and laser," in *Intelligent Vehicles Symposium (IV), 2015 IEEE*, June 2015, pp. 1–6.