A probabilistic distribution approach for the classification of urban roads in complex environments

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

A probabilistic distribution approach for the classification of urban roads in complex environments

Giovani B. Vitor[1,2], Alessandro C. Victorino[1] and Janito V. Ferreira[2]

***Abstract*— Navigation in urban environments has been receiving considerable attention over the past few years, especially for self-driving cars. Road detection for Autonomous Systems, and also for ADAS (Advanced Driving Assistance Systems) remains a major challenging in inner-city scenarios motivated by the high complexity in scene layout with unmarked or weakly marked roads and poor lightning conditions. This paper introduces a novel method that creates a classifier based on a set of probability distribution. The classifier, created using a Joint Boosting algorithm, aims at detecting semantic information in roads. This approach is composed of a set of parallel processes to calculate the superpixel using the Watershed Transform method and the construction of feature maps based on Textons and Disptons. As a result, a set of probability distribution is generated. It will be used as an input to model the week classifier by our Joint Boosting algorithm. The experimental results using the Urban-Kitty benchmark are comparable to the state-of-the-art approaches and can largely improve the effectiveness of the detection in several conditions.**

***Index Terms*— Road Detection, Computer Vision, Joint Boosting, Texton Map, Dispton Map, Watershed Transform.**

I. Introduction

Autonomous Navigation for urban environments has been receiving considerable attention from the robotic community over the past few years, motivating researchers to propose approaches towards the detection of roads in challenging inner-city environments. The inner-city road detection usually helps an Intelligent Vehicle System to get a better understanding of the environment improving interlinked or dependent tasks such as path planning [1], road following [2], and visual servoing [3].

Applications for road detection using camera sensors must deal with a set of problems such as: (i) continuously changing backgrounds in different environments (inner-city, highway, off-road), (ii) different road types (shape and color), (iii) the presence of different objects (signs, vehicles, pedestrian) and (iv) differences in imaging conditions (variation of illumination and weather conditions).

Many researchers have addressed this problem using monocular or stereo vision [4][5]. Approaches using monocular vision aim at detecting lane marking [6], appearance cues [7] or the 3D aspect by using prior knowledge about the environment as an extra source of information [8]. The detection methods using lane marking approaches may fail in unmarked roads, and some approaches overcome this assumption discriminating the overall road area with some appearance cues such as color feature [7][9][10], and texture feature [11][12]. 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. Detection using prior knowledge of 3D aspect of the road may fail due to the car's dynamic and road's imperfections. In the case of stereo vision, the 3D information of the environment is typically used to estimate free spaces and obstacles, with specific techniques like the V-Disparity Map [13][14]. However , the detection of road in urban environments must deal with different sources of noise that makes it difficult to define the disparity map. Some approaches propose improvements to the model by merging color, texture and adding 3D information [15][16].

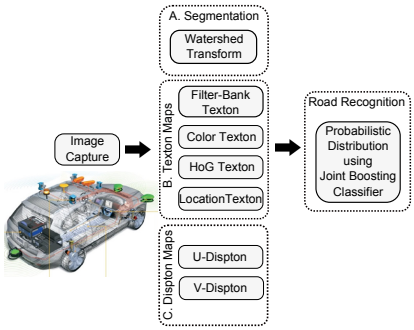

Fig. 1. Solution block diagram.

This work proposes a novel method for road detection in inner-city, using a set of probabilistic distribution to model the classifier of a Joint Boosting algorithm. Differently from others works, this approach creates a different set of features merging a technique called Diston, proposed by previous works (3D information) with the Texton (2D texture and color) to compute a set of probabilistic distribution for each superpixel. The probabilistic distribution feature's descriptor is used to model the weak classifier used in the Joint Boosting algorithm.

The rest of the paper is organized as follows: Section II presents the Image Segmentation, Texton Mapping and Dispton Mapping processes. Section III presents the formalism to model a set of probabilistic distribution using a Joint Boosting algorithm. Section IV presents the results of an experiment using the KITTI benchmark. Finally, Section V presents some conclusions about what we have learned from the study and proposals for future work.

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

II. Image Processing

In this section is presented a description of three parallels processes namely Image Segmentation, Texton Map and Dispton Map. As can be seen in Figure 1, the resulting output of these processes will be used as a source for generating the set of probability distribution in a second step.

A. Image Segmentation

Among various approaches used to produce superpixels, the most used nowadays in literature is the mean-shift algorithm [17][18]. In this work however, it is explored another methodology based on Watershed Transform that since [19] has been applied to road detection. The combination of Watershed Transform with other filters has presented encouraging results as shown in [20].

In order to obtain a reasonable flexibility to determine the segmentation level, we have used the same approach described in [20]. Three pre-filters were added: (i) the *Morphological Gradient Adjusted* (MG_{adj}), (ii) the *AreaClose* and (iii) the *Hmin*. Filter (i) is applied to obtain the high frequency image, where an adjustment is done to improve the low-contrast of high frequency in shadow areas [21]. The formal definiton can be seen in (1):

$$MG_{Adj} = \begin{cases} c\left[(f \oplus g_e) - (f \ominus g_i)\right]^{\gamma} & \text{, if } \{\forall x | f(x) < \rho\} \\ (f \oplus g_e) - (f \ominus g_i) & \text{, otherwise} \end{cases} \quad (1)$$

In equation (1) f is the image function, g_e and g_i are structuring elements centered at the origin, the operators $\oplus$ and $\ominus$ are respectively dilation and erosion. The non-linear transformation at low-contrast of high frequency has the factor γ setted to 0.45 and ρ is the threshold. The constant of normalization c is defined by the equation (2).

$$c = \frac{max((f \oplus g_e) - (f \ominus g_i))}{max((f \oplus g_e) - (f \ominus g_i))^{\gamma}} \quad (2)$$

The goal of the filters (ii) and (iii) is to control the segmentation level of the Watershed Transform by acting on the regional minimum of GM_{Adj}. The procedure is identical to [20], where the parameter λ of *AreaClose* determines the area of the regional minimum to be cut out, and the parameter h of *Hmin* determines the height from the regional minimum to be also cut out. The final outcome applied in a sample image can be observed in Figure 2. Notice that the classification's procedure is performed making a probability distribution for each super-pixel. Therefore, the analysis of these parameters is done to understand their sensibility in the final result of the segmentation.

B. Texton Map

In the last decade, *Textons* have been proven effective for generic feature representation of object [18][22], where a class demands different appearances to have a compact representation maintaining their efficiency. Thereby, the methodology applied for this block is to learn a dictionary of Textons using a textonization technique [23], which allows to perform a dense-texture-based feature extraction for all pixels. The

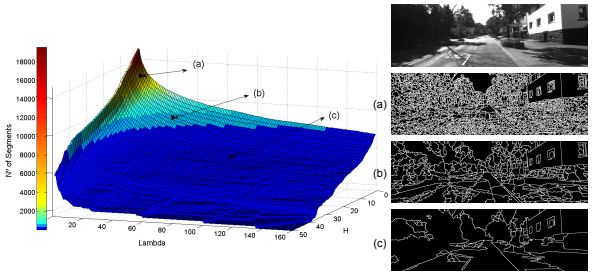

Fig. 2. Example of an influence surface for the parameters λ and h on the number of segments in an given image. In (a) $\lambda_1 = 5$, $h_1 = 2$ and 4090 segments; (b) $\lambda_2 = 30$, $h_2 = 5$ and 427 segments; (c) $\lambda_3 = 80$, $h_3 = 15$ and 73 segments.

process of *textonization* generates the *Texton Map*, having the same size of the image. The Textons contained in the dictionary have their value associated with all pixels in this map. It can be seen as a pre-classification or a transformation from feature's space to the texton's space. Thus, this process is done by applying the K-Means algorithm on a feature's space. Denoting a dictionary as D, each texton's element $x_j \in D = \{x_1, x_2, ..., x_K\}$ represents a cluster generated by the algorithm, employing the Euclidian-distance as a metric. Finally, it is obtained the *Texton Map* $T \in \mathbb{N}^2$ with the pixel i having value $x_j \in D$.

In this work, the *textonization* executed in various feature's spaces, as it can be seen in [18]. The set includes 17-dimensional filter bank, 3-dimensional CIELAB color, 81-dimensional histograms of oriented gradient [24] 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 assigned to $D^b = 400$ clusters, $D^c = 128$ clusters,$D^g = 150$ clusters and $D^l = 144$ clusters, respectively. The output result for this module can be seen in Figure 3.

C. Dispton Map

Based on the approach explained in section II-B, *Texton Maps* are able to discriminate between class of similar textures. However this technique lacks spacial information. This section presents an approach to build two additional dictionaries over 3D information from Stereo Vision. This method is called *Dispton Map* and it aims at creating meaningful clusters based on the Disparity Map, denoted by I_Δ. Attempting to have the same functional advantage provided by the usage of U-Disparity and V-Disparity algorithm to filter and extract the navigable area and obstacles in literature [13][25], this work addresses another way to embed these information in a dictionary of Dispton, generating the *Dispton Maps* from I_Δ.

Firstly, the technique consists in putting in evidence the peaks of the U-V Disparity maps, which concentrates the relevant information to start the process of *Disptonization*. Defining U-Disparity as $I_u\Delta$ and V-Disparity as $I_v\Delta$, they are obtained from a histogram for each column, $I_u\Delta = \{hist(I\Delta(:,u))|\forall u \in \{0..width-1\}\}$ and for each row,

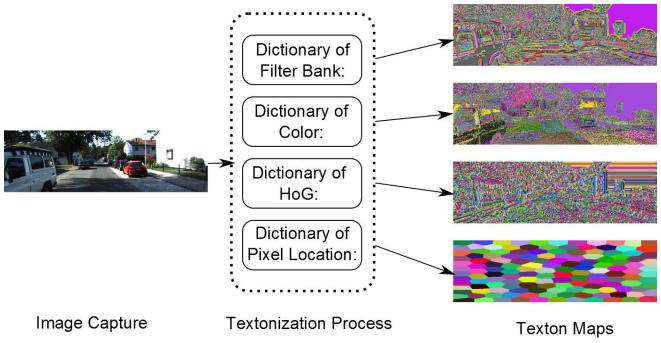

Fig. 3. The texton maps resulting from the textonization process using different features.

$I_v\Delta = \{hist(I\Delta(v,:))|\forall v \in \{0..height-1\}\}$. Like Watershed Transform, they can be seen as surfaces to apply the *Hmin* and filter the regional minimum (in this case considered as noise) of these surfaces. The result maps, denoted $I_u^h\Delta$ and $I_v^h\Delta$ are calculated by the binarization. A Hough Transform is executed to detect line segments, characterized by l^u and l^v.

After that, to build the dictionary of *U-Dispton* (D^u) is applied the clusterization where the points from each line segment l_j^u , supplies seeds to perform the clusterization of dispton's element $j \in \{1..NumberOfLines\}$. In equation (3), the clusterization process of a line segment l_j^u, denoted by $\Lambda(I_u^h\Delta)(l_j^u)$, is given by:

$$\Lambda(I_\eta^h\Delta)(l_j^\eta) = \begin{cases} C_j^\eta & \text{,if } \{x\} \in N_x(I_\eta^h\Delta) \neq 0 \\ 0 & \text{, otherwise} \end{cases} \quad (3)$$

Where the cluster C_j^η is defined in equation (4). Note that the variable (η) can be either u or v.

$$C_j^\eta = \{j|x_1W + x_2 \in \{l_j^\eta\} \subset N_x(I_\eta^h\Delta)\} \quad (4)$$

Here, W is the number of image columns, and x_1 and x_2 are the row and column coordinates respectively, and the term $N_x(.)$ represents the neighbors of the x element. There are two additional clusters given by $C_1^u = \{I_u\Delta(v,:)|v \in \{0,1,..,\tau\}\}$ and $C_2^u = (I_u\Delta \cap C_1^u)'$, where ($'$) represents its complement. Therefore, the D^u is given by the union of all clusters (5):

$$D^u = \{ \bigcup_{\forall j \in l^u} \Lambda(I_u^h\Delta)(l_j^u) \cup C_1^u \cup C_2^u \} \quad (5)$$

Finally, with the *U-Dispton* dictionary, it is possible to obtain the *U-Dispton map* (Dm_u) by the following equation (6):

$$Dm_\eta = \{D^\eta(I\Delta(x))|\forall x; I\Delta(x) \neq 0\} \quad (6)$$

In order to build the *V-Dispton* dictionary (D^v), the clusters are created separately. Following Hu and Uchimura [26], a road is modeled as a plane so that it can be represented by straight slope line segments in the V-disparity map. In this sense, the goal of the first cluster is to curve fitting these line segments to represent the surface of navigable area. From l^v set, a subset l^{vs} is obtained filtering out the line segments with a given vertical orientation (7):

$$l^{vs} = \{l_i^v|\forall l_i^v; Ang(l_i^v) < 90° - \psi\} \quad (7)$$

The $Ang(.)$ represents the angle of inclination with the reference defined on the bottom-left image and ψ a parameter of input. Thus, the road surface can be formed by a succession of plane's parts, being projected as a piecewise linear curve [27]. In order to connect the line segments that represent the surface, the algorithm sorts the l^{vs} set based on the distance from the line segment to the reference. Starting from the first line segment l_0^{vs} to the last one l_n^{vs}, the constraints that define whether two line segments can be connected, are given by the follow equation (8), which l^{vc} is the set of connected line segments.

$$l^{vc} = min(dist(l_i^{vs}, l_j^{vs})) \begin{cases} \forall l_j^{vs} \in \{l^{vs} > l_i^{vs}\} \text{ and} \\ \text{if } l_j^{vs} \subset AreaSupport(l_i^{s-}, l_i^{s+}) \end{cases} \quad (8)$$

Where the function $dist(.)$ between two line segments is calculated considering the Euclidian distance from the nearest points of the current line segments, limited by a given maximum distance ε between them. The function $AreaSupport(.)$ delimits the search area by two line segments as seen in (9).

$$AreaSupport(l_{s1}, l_{s2}) = \begin{cases} 1 & \text{,if } right(l, l_{s1}) \text{ and } left(l, l_{s2}) \\ 0 & \text{,otherwise} \end{cases} \quad (9)$$

This area is defined by a translation from l_i^{vs} given by the σ parameter, then $l_i^{s-} = l_i^{vs} - \sigma$ and $l_i^{s+} = l_i^{vs} + \sigma$. The other two functions in this equation return true case when the line segment is on right and left of the reference lines. As a result, the cluster C_1^v is obtained applying the equation (3) on the l^{vc} set ($\Lambda(I_v^h\Delta)(l_j^{vc})$) with one more constraint, where all pixels cannot cross out the line (l_{lim1}) formed by the first and last points of l^{vc} (added a small shift constraint). In addiction, the second cluster is generated taking those pixels which cross out the first one and is restricted to another shifted line $l_{lim2} = l_{lim1} + \sigma_2$, resulting the equation (10):

$$C_2^v = \{\Lambda(I_v^h\Delta)(l_j^{vc})| \text{ if } AreaSupport(l_{lim1}, l_{lim2})\} \quad (10)$$

To finish the *Disptonization*, the last two clusters are generated by $C_3^v = \{I_v\Delta(:,u)|\forall u \in \{1,..,\tau\}\}$, where τ defines the max disparity to be considered as background or infinity, and $C_4^v = \{(I_v\Delta \cap (C_1^v \cup C_2^v \cup C_3^v))'\}$. The *V-Dispton* dictionary (D^v) is defined as:

$$D^v = \{C_1^v \cup C_2^v \cup C_3^v \cup C_4^v\} \quad (11)$$

And the generation of *V-Diston map* is obtained by the equation (6). Algorithm 1 summarizes the *Disptonization* process.

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.* for l^u and l^v;
4: Determine the *U-Dispton* dictionary D^u by eq. (5):
- Apply the clusterization on l^u, eq. (3);
5: Determine the *V-Dispton* dictionary D^v by eq. (11):
- Filter out the vertical lines to take l^{vs}, eq. (7);
- Find out the connected lines l^{vc}, eq. (8);
- Define the clusterization to $C_1^v, C_2^v, C_3^v, C_4^v$;
6: Generate the *UV-Dispton map* by eq. (6)

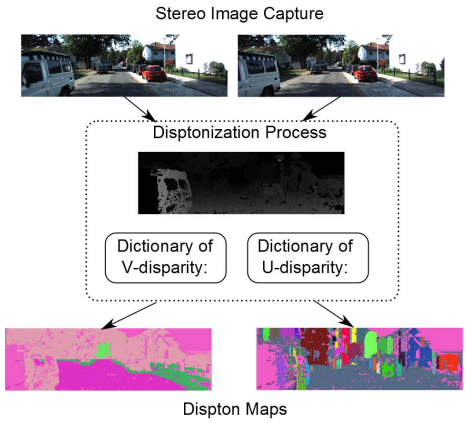

Fig. 4. The Dispton maps obtained from the disptonization process using the Disparity Map.

Note that the *V-Dispton map* has 4-dimensional clusters and the *U-Dispton map* has N-dimensional clusters. Intuitively, they aim at storing important information such as navigable area, sidewalk, obstacles and background. The N-dimensional structure from *U-Dispton map* dynamically retrieve the representation of all possible different obstacles in the scene, as can be seen in Figure 4. With the *Texton maps* and *Dispton maps*, the next section explains how they are combined with the superpixel to perform the classification.

III. Road Recognition

This section presents an approach to represent and compute the classification of the road class, as shown in previous sections, where the road recognition can be executed using an adapted version of the Joint Boosting algorithm [28]. In fact, the algorithm is inspired by the *TextonBoost* approach [29], which iteratively builds a *strong classifier* as a sum of *week classifiers*, simultaneously selecting discriminative features. We have improved the representation of weak classifiers using a specific shape filter. Thus, the novelty is to build a set of probability distribution of the Texton and Dispton maps from the decomposition of the scene into a number of semantically consistent regions, supplied by the segmentation result shown in section II-A, to model the *weak classifier*.

The process could be formally explained taking into account the maps $\{M^f : f \in \{F\}\}$ where $F = \{b, c, g, l, v, u\}$ is the set of Textons and Disptons. Each element i in the map $M^f \in \mathbb{N}^2$ belongs to exactly one region, identified by its region-correspondence variable $S_r \in \{1, ..., NumSegments\}$. The r-th region is then simply the set of elements i_r whose region-correspondence variable equals r, i.e., $i_r = \{i : M_i^f = r\}$. We use $X_i^f = \{X_1^f, X_2^f, ..., X_N^f\}$ to denote the set of random variables corresponding to the f-th value of i-th element into M^f. Any possible assignment to the random variables $X_i^f = x_j^f$ takes values from $j \in D^f$, which D^f is defined by the constructed dictionary for each $f \in F$ generated in the sections II-B and II-C, .

The probability of the X_i^f if given by $P(X_i^f = x_j^f)$, and the associated set of probability distribution under the S_r is denoted by $P(X_r)$, as can be seen in the equation (12):

$$P(X_r) = \{ \bigcup_{f \in F} \left\{ \frac{1}{Z} \sum_{i_r} P(x_j^f) \right\} | \forall j \in D^f \} \tag{12}$$

In Equation (12), Z is a normalization factor for each probability distribution set. Using the probability representation of Textons and Disptons, the *weak classifiers* are modeled as comparisons of this probability distribution to a decision stump based on a threshold, where each *weak classifier* is *shared* between a set of classes, allowing a single probability to help classify several classes at once. They are defined by wc containing 2-tuples $[x_{rand}, P(x_{random})]$, where the first component represents a random possible assignment $\{x_{rand} : x_j^f \in D^f\}$ and its value of probability randomly defined. To express how well the probability distribution of $P(X_r)$ at a given x_j^f matches the weak classifier, a comparison response is given by equation (13):

$$d(wc, S_r) = 1 - \sqrt{[P(x_{rand}) - P(X_r = x_j^f)]^2} \tag{13}$$

Thereby, the Joint Boosting algorithm is an additive model of the form $H(c_l) = \sum_{m=1}^{M} h_m(c_l)$, that sum the classification confidence of M joint weak classifiers. In this case, $H(c_l)$ represents the strong learned classifier and the weak classifiers are extended to discriminate the share between classes. Therefore, each weak-learner is modeled as a decision stump of the form:

$$h(c_l) = \begin{cases} a\delta(d(wc, S_r) > \theta) + b & \text{, if } \{c_l \in L\} \\ \kappa_{c_l} & \text{, otherwise} \end{cases} \tag{14}$$

Where $\delta(.)$ is a 0-1 indicator function. The share is given by those classes ($c_l \in L$), where the weak learner gives $h(c_l) \in \{a+b, b\}$ depending on the comparison of $d(wc, S_r)$ to a threshold θ. The constant k_{c_l} ensures asymmetrical sets of positive and negative training examples for those classes that do not share the feature ($c_l \notin L$). Thus, the resulting classification output is defined by the probability conversion given by(15):

$$P = \frac{1}{Z} \exp^{-H(c_l)} \tag{15}$$

Here, the Z represents the normalization factor into the classes $c_l \in L$.

IV. Experimental Results

In this subsection we present the results of our experiments using real driving situations. It is used the Urban Kitti-road dataset[1], which consists of $\simeq$ 600 frames (375x1242 px) recording from five different days and containing relatively low traffic density [30]. The data are categorized in three sets having each one a subset of training and test images, representing a typical road scene in inner-city. The first set called the UU is formed by images taken from an urban unmarked area and has 98 images for training and 100 images for testing. The second set called the UM is formed by images taken from an urban marked two-way road and has 95 images for training and 96 images for testing. Finally, the final set called UMM is formed by images taken from an urban marked multi-lane road and has 96 images for training and 94 images for testing. The experiments in this work use the training set in the perspective space to learn the classifier and the metric space to calculate the complete evaluation of this approach. The evaluation process is done on the metric space in order to capture the fact that vehicle control happens in the 2D environment. Further, the evaluation in perspective space is biased by the fact that the pixel's value in near range is more homogenous and covers a larger area of the evaluated perspective pixels [30].

The learning process were executed separately for each category. 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. Table I shows the results of the quantitative evaluation of our approach applied in the test set. For comparison purposes, it also presents the baseline provided as a lower bound, by averaging all ground truth road maps from the present testing set, and also the results of the proposed road detection using Artificial Neural Network (ANN) [20]. As can be seen for the UU category, our approach reached an improvement of 26.12% if compared to ANN and 10.80% to Baseline, using the testing set on the metric space. With respect to the UM category, our approach reached a level of correctness of 87.60%, reaching a gain of 24.96% compared to ANN and overcoming the baseline approach in 5.07%. Using the UMM category, which is less complex if compared to others, we can highlight our approach, overcoming the ANN approach with 9.03% and 13.95% compared to the baseline. The qualitative result for this challenging dataset can be seen in Figure 5, presenting a classified image using the perspective space. The same images using the metric space can be seen in Figure 6.

To conclude the evaluation process, Table II presents the final results merging all categories. Our approach presents 87.21% of correctness for challenging urban Kitti-road benchmark. According to our experiments we believe that our approach improved the detection of roads in scenarios with high complexity, significantly outperforming the baseline.

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

TABLE I

Results [%] of pixel-based for the all categories on the metric space evaluation.

Urban Unmarked (UU)						
	F_{max}	AP	Prec.	Recall	FPR	FNR
Baseline [30]	*69.49*	**73.84**	*65.73*	73.70	*12.78*	26.30
ANN [20]	54.17	36.86	39.50	**86.19**	43.92	**13.81**
Our	**80.29**	*69.05*	**85.58**	*75.61*	**4.24**	*24.39*
Urban Marked (UM)						
	F_{max}	AP	Prec.	Recall	FPR	FNR
Baseline [30]	*82.53*	**85.59**	*79.24*	*86.11*	*10.41*	*13.89*
ANN [20]	62.64	46.80	50.18	83.34	38.21	16.66
Our	**87.60**	*76.04*	**85.92**	**89.36**	**6.76**	**10.64**
Urban Marked Multi-Lane (UMM)						
	F_{max}	AP	Prec.	Recall	FPR	FNR
Baseline [30]	76.17	*78.42*	65.02	91.95	57.89	8.05
ANN [20]	*81.09*	68.93	*70.43*	**95.56**	*46.94*	**4.44**
Our	**90.12**	**85.04**	**88.15**	*92.12*	**14.50**	*7.82*

Fig. 5. The resulting output of our approach in perspective space. The first, second and third rows show the UU, UM and UMM category image respectively.

TABLE II

Results [%] of pixel-based for complete URBAN ROAD area evaluation.

	F_{max}	AP	Prec.	Recall	FPR	FNR
Baseline [30]	*75.61*	**79.72**	*68.93*	83.73	*21.73*	16.27
ANN [20]	68.12	51.52	54.85	**89.85**	42.59	**10.15**
Our	**87.21**	*77.79*	**86.96**	*87.47*	**7.55**	*12.53*

V. Conclusions and Future Works

In this paper, we have proposed an approach for road recognition for inner-city based on appearance, shape and spacial model learned from training data. The main contribution of this work is the creation of a probabilistic distribution based on Texton and Dispton maps to model weak classifiers used in a Joint Boosting classifier.

Experiments conducted on real driving situations demonstrate the qualitative and quantitative evaluation of our algorithm to detect road despite the presence of shadows and other objects in the scene, inherent from the complexity of inner-city environments. The result also provides the benefits of our approach over existing methods.

We are still working on improvements to reduce the processing time using the GPU architecture, and also working

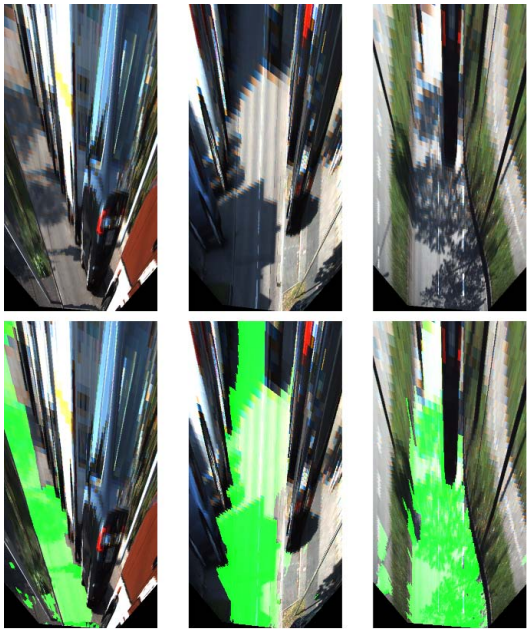

Fig. 6. The resulting output of our approach in metric space (highlighted in green). The first, second and third columns show the UU, UM and UMM category image respectively, from the images of Figure 5.

on Self-Organizing Maps [31], in order to better discriminate the road pattern and extending the recognition for different classes such as vehicles, builds, sidewalks, etc. This approach could improve the prediction of the classifier. The complete application will be embedded in a real car-like robot, sponsored by the project ROBOTEX, from the Heudiasyc laboratory, to perform autonomous driving in urban environments.

REFERENCES

[1] N. Rawashdeh and H. Jasim, "Mult-sensor input path planning for an autonomous ground vehicle," in *Mechatronics and its Applications (ISMA)*, 2013, pp. 1–6.

[2] C. Thorpe, M. Hebert, T. Kanade, and S. Shafer, "Vision and navigation for the carnegie-mellon navlab," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 3, pp. 362 – 373, May 1988.

[3] 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.

[4] J. Alvarez, A. Lopez, and R. Baldrich, "Illuminant-invariant model-based road segmentation," in *Intelligent Vehicles Symposium, 2008 IEEE*, june 2008, pp. 1175 –1180.

[5] F. Dornaika, J. Alvarez, A. Sappa, and A. Lopez, "A new framework for stereo sensor pose through road segmentation and registration," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 4, pp. 954 –966, dec. 2011.

[6] M. Felisa and P. Zani, "Robust monocular lane detection in urban environments," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*, June 2010, pp. 591–596.

[7] 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.

[8] P. Coulombeau and C. Laurgeau, "Vehicle yaw, pitch, roll and 3d lane shape recovery by vision," in *Intelligent Vehicle Symposium, 2002. IEEE*, vol. 2, June 2002, pp. 619–625 vol.2.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] A. Broggi, C. Caraffi, R. I. Fedriga, and P. Grisleri, "Obstacle detection with stereo vision for off-road vehicle navigation," in *Proceedings of the International IEEE Workshop on Machine Vision for Intelligent Vehicles*, 2005, pp. 1–8.

[15] 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.

[16] 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.

[17] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 5, pp. 603–619, 2002.

[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] S. Beucher and M. Bilodeau, "Road segmentation and obstacle detection by a fast watershed transformation," in *Intelligent Vehicles '94 Symposium, Proceedings of the*, 1994, pp. 296–301.

[20] 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.

[21] 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.

[22] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," *CoRR*, vol. abs/1210.5644, 2012.

[23] 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.

[24] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *In CVPR*, 2005, pp. 886–893.

[25] N. Soquet, M. Perrollaz, R. Labayrade, and D. Aubert, "Free Space Estimation for Autonomous Navigation," in *5th International Conference on Computer Vision Systems*, Bielefeld, Allemagne, 2007. [Online]. Available: http://hal.inria.fr/hal-00780658

[26] Z. Hu and K. Uchimura, "U-v-disparity: an efficient algorithm for stereo vision based scene analysis." in *Intelligent Vehicles Symposium (IV),IEEE*, 2005, pp. 48–54.

[27] N. Hautiere, R. Labayrade, M. Perrollaz, and D. Aubert, "Road scene analysis by stereovision: a robust and quasi-dense approach," in *Control, Automation, Robotics and Vision, 2006. ICARCV '06. 9th International Conference on*, 2006, pp. 1–6.

[28] 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.

[29] 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.

[30] 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.

[31] T. Kohonen, *Self-organizing maps*, 3rd ed., ser. Springer series in information sciences, 30. Berlin: Springer, December 2001.