



HAL
open science

A Histogram-Based Joint Boosting Classification for determining Urban Road

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

► **To cite this version:**

Giovani Bernardes Vitor, Alessandro C. Victorino, Janito V. Ferreira. A Histogram-Based Joint Boosting Classification for determining Urban Road. 17th IEEE International Conference on Intelligent Transportation Systems (ITSC 2014), Oct 2014, Qingdao, China. pp.2245-2246. hal-01089100

HAL Id: hal-01089100

<https://hal.science/hal-01089100>

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.

A Histogram-Based Joint Boosting Classification for determining Urban Road (Extended Abstract)

Giovani B. Vitor^{1,2}, Alessandro C. Victorino¹ and Janito V. Ferreira²

Abstract—Road detection for inner-city scenarios remains a difficult problem due to the high complexity in scene layout with unmarked or weakly marked roads and poor lighting conditions. This paper introduces a novel method based on multi normalized-histogram with Joint Boosting algorithm to road recognition. The approach performs three modules in parallel that are the Image Segmentation, the Texton Maps and the new one Dispton Maps. The first one applies a combination of pre-filters with Watershed Transform to make the super-pixel. The last two perform a dense feature extraction based on 2D texture image and 3D disparity image to get appearance, shape and context information. At last, a discriminative model of road class is learned based on distribution of Textons and Disptons applied in Joint Boosting algorithm. The proposed work reports real experiments carried out in a challenging urban environment utilizing the modern KITTI benchmark for road areas in which meaningful evaluation can be done to illustrate the validity and application of this approach.

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

I. INTRODUCTION

Road recognition is a key for autonomous driving systems, such as path planning, road following, and navigable tasks such as visual servoing and line detection. From the last years, several approaches related to vision-based road detection have been proposed, varying from mono to stereo vision sources. They aim to solve several challenging problems such as the continuously changing backgrounds in different environments, different road types, the presence of different objects and also different imaging conditions. In general, the approaches for road detection using Mono vision are based on color, texture, or the combination of both [1], [2], [3]. Despite the rich amount of information present in the color feature, which imposes less physical restrictions by performing the detection according to some similarity, it may fail due to high intra-class variability present in the dynamic nature of the scenes. The texture is scale-dependent and is affected by the strong perspective in road image. The V-disparity Map [4] technique is an example of approaches that use 3D information as a source. However, these techniques must deal with different noise sources, such as shadow, road texture, light variations, that become hard to create the disparity map. An alternative to overcome these constraints is to combine 2D and 3D data, enhancing

This work was supported in part by the French Picardie project VERVE and the European Fund of Regional Development FEDER. Authors are members of ¹Heudiasyc UMR CNRS 7253 Université de Technologie de Compiègne, ²Universidade Estadual de Campinas (UNICAMP). Giovani B. Vitor holds a Ph.D. scholarship from CNPQ. Contact authors giovani.bernardes-vitor@hds.utc.fr

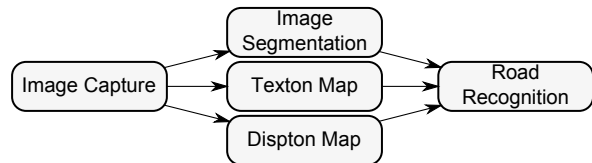


Fig. 1. Solution block diagram.

the detection and reducing the drawback of each approach running separately [5].

This work proposes a novel method for road recognition in inner-city scenarios, merging 2D and 3D information, resulting in a multi normalized-histogram that model the appearance, shape and context using a Joint Boosting algorithm. Differently from others works, this approach create a novel form to represent the 3D information encapsulated on *Disptons* to compose jointly with the popular *Textons*, a different set of features, to better represent the road class given by the complex environment. A discriminative model for road class is learned based on the multi normalized-histogram performed on super-pixel and classified using a Joint Boosting classifier.

II. MULTI HISTOGRAM-BASED JOINT BOOSTING

This section presents a brief understanding of the proposed procedure, as shown in Figure 1. The Image Segmentation block is inspired by [5], which performs a combination of pre-filters such as the *gradient of the image*, the *AreaClose* and the *Hmin*, jointly with the Watershed Transform to get the super-pixel.

The Texton Map block implements the Texton method that applies the textonization process based on the extraction of various features to get the *texton maps* [6]. 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. This work uses the same set of feature's descriptor proposed by [7]. In the same way of the 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_Δ). This technique is called by *Disptonization*. It encapsulates 3D information provided by the U-Disparity ($I_{u\Delta}$) and V-Disparity ($I_{v\Delta}$) in a dictionary of Dispton, taking then the *Dispton Maps* from I_Δ . The summarized Disptonization algorithm can be seen in Algorithm (1).

Algorithm 1 Disptonization algorithm:

- 1: Process $I_{u\Delta}$ and $I_{v\Delta}$ from I_Δ ;
- 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 and the V -Dispton dictionaries;
- 5: Generate the UV -Dispton maps;

The Road Recognition block uses a classification method derived from *TextonBoost* [6] to model the road class. The proposed multi normalized-histogram joint boosting algorithm builds a *strong classifier* as a sum of *weak classifiers*. Based on super-pixels, the 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. 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 χ^2 metric comparisons of the weak classifiers' histogram with the response's histogram obtained from the super-pixels of 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 super-pixel i . The $d(\cdot)$ is the χ^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 θ . The constant κ_{c_i} ensures asymmetrical values of positive and negative training examples.

III. EXPERIMENTAL RESULTS

The proposed algorithm is experimentally validated using the URBAN KITTI-ROAD dataset. The data are categorized in three sets, called UU, UM and UMM, having each one a subset of training and test images. The experiments use the training set in the perspective space to learn the classifier and the metric space to calculate the complete evaluation of this approach, as suggested by [8].

The learning processes were executed separately for each category. Table I presents the final results merging all these categories. Our approach presents 83.41% of this challenging urban Kitti-Road benchmark, reaching a gain of 7.8% from BL, 15.29% from ANN and 4.66% from SP algorithm. This improvement is significative considering the complex environment where the system is exposed and seen that a small set was used for training, reaching a desired generalization.

Some image samples collected from the test set can be seen in Figure 2. From the qualitative analysis, it is possible to see that the method proposed can solve the problem imposed by different obstacles such as vehicles, trees and

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

	F_{max}	AP	Prec.	Recall	FPR	FNR
Baseline [8]	75.61	79.72	68.93	83.73	21.73	16.27
ANN [5]	68.12	51.52	54.85	89.85	42.59	10.15
SP [9]	78.75	77.66	72.41	86.30	18.93	13.70
Our	83.41	74.06	82.39	84.46	10.39	15.54

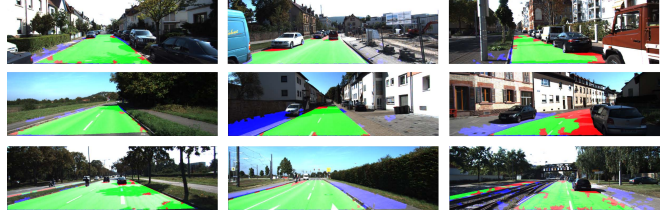


Fig. 2. Qualitative result images for challenging scenarios. The lines represent the output result for the UU, UM and UMM categories respectively. Red color denotes false positives, blue color areas correspond to false negatives and green color represents true positives.

sidewalks. Even with different conditions such as shadow areas and areas with high luminosity provided by the sun, the algorithm can acquire satisfactory approximations for the road area.

IV. CONCLUSIONS

The work proposed in this paper provides an approach for road recognition for inner-city scenarios based on the novel method that join a multi normalized-histogram with a Joint Boosting algorithm. The main advantages of the method are: (i) Its independence from prior knowledge of road assumptions; (ii) The integration of stereo vision not only to improve detection precision but also to extend the recognition capability on different urban elements; (iii) Module independency leverages the usage of parallelization and improvement on resources usage.

REFERENCES

- [1] 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.
- [2] 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.
- [3] 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.
- [4] A. Broggi, C. Caraffi, R. I. Fedriga, and P. Grisleri, "Obstacle detection with stereo vision for off-road vehicle navigation," in *Proc. of the IEEE Workshop on Machine Vision for Intelligent Vehicles*, 2005, pp. 1–8.
- [5] 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.
- [6] 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," *IJCV*, pp. 2–23, 2009.
- [7] L. Ladicky, C. Russell, P. Kohli, and P. H. S. Torr, "Associative hierarchical crf's for object class image segmentation," in *Computer Vision, IEEE 12th International Conference*, 2009, pp. 739–746.
- [8] 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.
- [9] N. Einecke and J. Eggert, "Block-matching stereo with relaxed fronto-parallel assumption," in *IV*, 2014.