# Towards automatic assessment and mapping of traffic infrastructure by adding vision capabilities to a geoinformation inventory

S. Šegvić, K. Brkić and Z. Kalafatić
Fac. of Electrical Engineering and Computing
Unska 3, 10000 Zagreb
e-mail: name.surname@fer.hr

V. Stanisavljević
Tilda d.o.o.
Medvedgradska 1, 10000 Zagreb
Email: vladimir@tilda.hr

D. Budimir and I. Dadić
Inst. of Traffic and Communications
Kušlanova 2, 10000 Zagreb
Email: budimir@ipv-zg.hr

***Abstract*—We consider geoinformation inventory systems containing locations and types of traffic signs and road surface markings in a given geographical region. Such systems allow a human operator to assess the current state of the traffic infrastructure by comparing a recent georeferenced video with the prescribed state stored in the inventory. We are concerned with automating this assessment procedure (as well as the initial creation of the inventory) by designing appropriate computer vision modules. In this paper, we briefly present our research project, provide a review of suitable computer vision techniques, and present in more details two techniques which proved most promising in preliminary experiments. The two presented techniques are traffic sign detection by boosted Haar wavelets, and enhancement of surface markings by applying steerable filters to suitably preprocessed images. The experiments have been performed on individual real images acquired from a moving vehicle by a single perspective camera. The obtained results are provided and discussed.**

## I. Introduction

Traffic safety is an important issue in our society [1], which critically depends on compliance of traffic control devices with relevant regulations. The compliance needs to be periodically assessed in order to be able to detect hazardous situations due to stolen, broken, erased, removed or otherwise inadequate devices (traffic signs and surface markings). The assessment responsibility is typically assigned to local authorities with limited resources, which usually delegate the implementation to external contractors [2]. Thus, the assessment frequency and the resulting safety standards directly depend on the efficiency of the assessment procedure.

We consider an effective and affordable technology to achieve high compliance standards of road and railway traffic control devices. The technology is based on (i) locating the elements of the traffic control infrastructure in a georeferenced video, and (ii) storing the extracted information in a geoinformation inventory system [3], [4], [5], [6]. Georeferenced video is acquired by synchronizing acquisition of image frames with the readings obtained from a GPS receiver. In order to increase the accuracy, the GPS localization is usually combined with relative localization techniques [7] such as odometry and inertial sensors [3], [5]. Once the inventory with the prescribed state of the traffic infrastructure is created, the assessment becomes a significantly simpler task. It suffices to acquire a recent georeferenced video and check the presence and visibility of all elements from the inventory. Thus, the creation of the traffic inventory, which is often referred to as *mapping*, is in principle performed only once, while the *assessment* is performed periodically, several times per year, depending on the available funding. Geometric relations during the assessment are depicted in Figure 1.

This research has been jointly funded by Croatian National Foundation for Science, Higher Education and Technological Development, and Institute of Traffic and Communications, under programme Partnership in Basic research, project number #04/20.

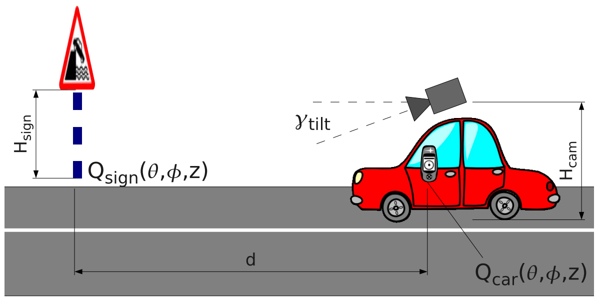



Fig. 1. Geometric relations between the camera, car, and the assessed road section ($Q_{\mathrm{sign}}$ is recovered during mapping, $Q_{\mathrm{car}}$ is measured).

Both the mapping and the assessment stage of the considered technology require that the elements of the traffic infrastructure are somehow located in the input video. Since performing that by hand is a tedious task, automatization approaches based on computer vision were evaluated by several previous research projects. A system presented in [5] focuses on accurate automatic mapping of traffic signs in stereo images acquired from a moving vehicle. The spatial displacement between the GPS receiver and the calibrated stereo rig has been recovered by a special calibration setup featuring a known 3D structure. Similar two systems based on monocular vision have been presented in [3], [6]. A slightly different approach relies on a handheld device incorporating a GPS receiver and a camera [4], [8]. Such approach, where the operator carries the sensing and computing device in her hands, is best suited for crowded areas in city centres.

In this paper, we present our current activities on a research project[1] aimed at providing automatic assessment and mapping capabilities to a geoinformation based traffic inventory system. The main goal of the project is to design effective computer vision techniques for detecting the traffic control infrastructure. In particular, here we focus on (i) detection and mapping of surface markings and curbs, and (ii) detection and recognition of traffic signs. The obtained results are eventually going to be implemented as components of an existing geoinformation

[1]The project web page is at http://www.zemris.fer.hr/~ssegvic/mastif

inventory system [9] which currently makes no use of computer vision, but instead relies on manual detection by human operators. Compared to previous works, our current research contribution is in the fields of (i) mapping of surface markings, and (ii) detection of traffic signs. In the future work, we would like to reinforce the detection by temporal [10] and contextual [11] constraints, by using the cues obtained from camera motion and global context of the interpreted road scene, respectively.

The paper is structured as follows. Section II presents the existing geoinformation inventory system developed at the partner organization. The relevant computer vision techniques are reviewed in Section III. Sections IV and V present our preliminary results towards the detection of traffic signs and mapping of surface markings. The paper is concluded in Section VI by a short discussion of the obtained results and some directions for the future work.

## II. A GEOINFORMATION TRAFFIC INVENTORY

A geoinformation inventory for traffic infrastructure has been actively developed at the Institute of Transport and Communications (cf. http://www.ipv-zg.hr) as a software product named OptaGIS. The developed system has been a successful tool in providing commercial roar maintenance assessment service to local authorities all over Croatia since 2005. Compared to competing solutions, OptaGIS offers simpler assessment due to support of georeferenced video, better interoperability with other software (AutoCAD, etc), as well as better presentation capabilities due to support of different coordinate systems (Gauss-Krūger, WGS 84, etc).

As illustrated in Figure 2 (left), OptaGIS organizes the stored data in several layers, each of which can be configured as visible or not. The most important layers in our context are the traffic inventory itself and georeferenced video acquired from the driver's perspective. Georeferenced video enables (i) an additional visualization dimension of the mapped road, and (ii) a deferred and repeatable analysis of the traffic infrastructure, and thus provides a qualitative improvement over straight-forward on-site assessment. The traffic inventory contains all information about the traffic infrastructure including traffic signs, surface markings and curbs, as shown in Figure 2 (right). Traffic inventory is especially useful in combination with georeferenced video, since that ensures simple procedures for mapping and assessment as sketched in the Introduction. OptaGIS allows an associative look-up across different layers: it is possible to display the desired part of the video by clicking on the mapped sign from the inventory, and vice versa.

The vehicle used for acquisition of georeferenced video is shown in Figure 3. Besides a calibrated perspective camera, the vehicle is equipped with a differential GPS receiver, as well as with inertial positioning and distance measuring instruments. Readings of the three positioning sensors are fused using a Kalman filter.

Fig. 3. The acquisition vehicle from outside (left), and inside (right).

## III. REVIEW OF RELEVANT VISION TECHNIQUES

This section reviews computer vision techniques which are relevant in our context. We first present state of the art in the detection of road surface markings in III-A, and then in III-B turn our attention to the recognition of traffic signs. Finally, we discuss opportunities for future research in III-C.

### A. Detection of road surface markings

Techniques for detection of surface markings and subsequent lane position tracking are needed in a number of important applications. Firstly, they are an important ingredient in driver assistance systems [1], [12]. Here, the extracted lane position is used to detect lane departures

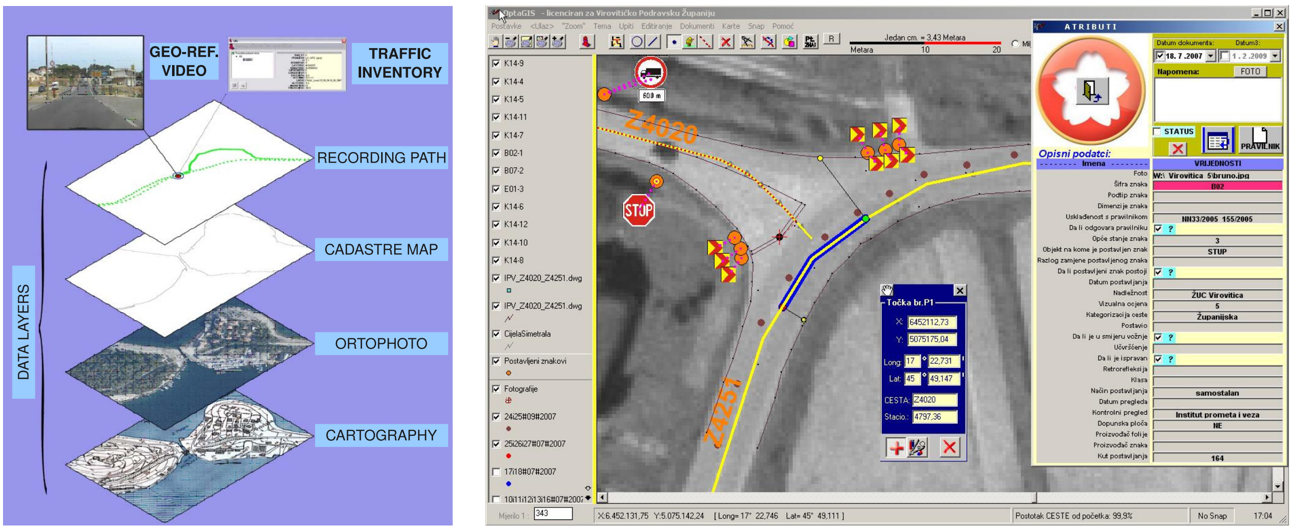



Fig. 2. Capabilities of the presented geoinformation system: some of the supported data layers (left), and a view of the traffic sign inventory layer shown above the ortophoto layer (right). Red circles in the right figure denote car positions during the mapping session.

and other dangerous conditions and accordingly warn the driver. Secondly, lane detection is useful in autonomous vision based navigation [13], although the current research is more directed towards unconstrained contexts [14]. Finally, lane detection can also be used for creating and assessing traffic inventories, which is our current application.

Most approaches for lane position tracking include some or all of the following elements [13], [12]:

- a dynamic model of the road and the vehicle
- a detection system for locating lane candidates in single images,
- a recursive filter capable for refining estimates over time, by assuming the vehicle and road models.

The road model is usually a low order polynomial (1-3). The model can be image based [15], [16], or ground-plane based [13], [12]. The latter model better interacts with the vehicle model, and more easily incorporates the parallelism of the lanes.

Many detection approaches have been used, but most of them are quite elementary and easy to implement. The approach from [13] utilizes horizontal search lines in the image plane to detect dark-light-dark grey-value transitions near the positions predicted by the road model. By assuming that lanes are the longest linear structures in the image, the lane directions can be detected by looking at peaks of a global gradient direction histogram [16]. An advantage of this approach is that it might be applied for locating curbs as well. However, detecting peaks in histograms can be tricky, especially in the presence of clutter. Long straight edges can be reliably detected by steerable filters [12]. The full lane model is consequently robustly fitted to the lane candidate pixels in a postprocessing step.

A technique called *inverse perspective mapping* (IPM) [17], [12] is especially suitable for lane detection, under a reasonable assumption of a locally planar road. The technique allows transforming source images taken from the driver's perspective into "inverse-perspective" images which would have been acquired by a camera aligned with the road surface normal. Inverse-perspective images are useful since the lane boundaries appear as parallel lines with constant width, as illustrated in Figure 4.

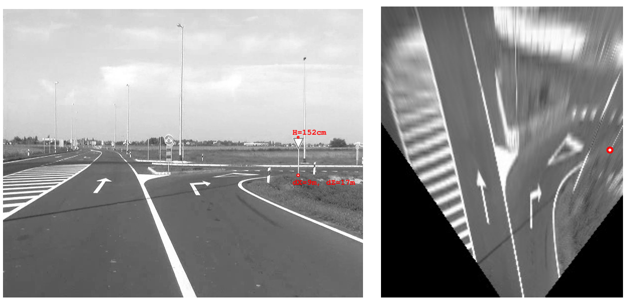


Fig. 4. Source image acquired from the driver's perspective (left) and the corresponding inverse-perspective image (right).

The existence of IPM is easy to prove. Denote points on the road as $\mathbf{q}_{\mathrm{R}}$, points in the perspective image as $\mathbf{q}_{\mathrm{P}}$, and points in the inverse-perspective image as $\mathbf{q}_{\mathrm{I}}$. In homogeneous coordinates, these points can be related by a 1:1 linear mapping known as homography [18]:

$$\mathbf{q}_{\mathrm{P}i} = \mathbf{H}_{\mathrm{RP}} \cdot \mathbf{q}_{\mathrm{R}i}, \forall i \tag{1}$$

$$\mathbf{q}_{\mathrm{I}i} = \mathbf{H}_{\mathrm{RI}} \cdot \mathbf{q}_{\mathrm{R}i}, \forall i \tag{2}$$

Thus we can see that, in the general case, IPM can be expressed by composing $\mathbf{H}_{RP}^{-1}$ and $\mathbf{H}_{RI}$:

$$\mathbf{q}_{Ii} = \mathbf{H}_{\mathrm{IPM}} \cdot \mathbf{q}_{\mathrm{P}i}, \forall i \tag{3}$$

$$\mathbf{H}_{\mathrm{IPM}} = \mathbf{H}_{\mathrm{RI}} \cdot \mathbf{H}_{\mathrm{RP}}^{-1} \,. \tag{4}$$

Parameters of the IPM matrix $\mathbf{H}_{\mathrm{IPM}}$ directly correspond to the geometry of the camera with respect to the road plane (cf. Figure 1). They can be calibrated beforehand by an adequate calibration procedure [19], [20]. In order to obtain best results, the calibration can be adapted to the dynamics of the vehicle motion by a suitable optimization procedure [21], [20].

### B. *Recognition of traffic signs*

State of the art sign recognition approaches usually employ a two-step procedure, consisting of *detection* and *classification*. The detection is concerned with locating a sign in the input image regardless of its identity, while the latter classification step identifies the exact type of the located sign.

Traffic sign detection approaches rely on one or more of the following cues: (i) geometric shape [22], [23], [24], (ii) colour, and (iii) appearance [25]. Shape-based detection can be based on Hough transform [23], or the so-called radial symmetry transform [22]. Colour-based detection relies on the specific colours (red, blue, etc) appearing within the the signs [25], [24]. Appearance-based approaches for object detection avoid the dependence on geometric models, usually by performing supervised learning on a set of representative samples of a desired object class. In particular, much success has been recently achieved by boosting Haar-like classifiers [26], [25]. A major advantage of such approaches is their applicability to objects which do not have an easily parameterizable geometric structure (e.g. faces [26]). However, the approach is not scale invariant, and therefore needs to be applied at several scales, which may result in misdetections [25].

The classification step has been less thoroughly addressed than detection. Most of the authors choose a classification approach without stating the reasons or comparing the results with some other classifier. Acceptable recognition results have been reported by using simple cross-correlation [22], cross-correlation with a trainable similarity measure [27], neural networks [25], [23] or support vector machines [24].

### C. *Open directions for further research*

Despite the considerable advances, a near-human recognition of traffic control devices has still not been reported. The methods fail due to spurious matches in the background, shadows or reflections, discretization issues and other moving objects in the scene [22], [25], [12]. We believe that significant improvement could be achieved by introducing additional constraints. These constraints are not easy to model, since we, humans, are so good in enforcing them, that we do not even have to consciously think about them. The constraints which in our opinion have not been fully exploited by previous researchers include smooth 2D motion [25], compliant 3D motion [10] and context [11]. Enforcing these constraints will eventually lead us to a unified approach in which the detection of the lane boundaries will reinforce the recognition of traffic signs.

## IV. Some results in detection of road surface markings

As stated in III-A, the detection of road surface markings is considerably simpler in inverse-perspective images. IPM is particularly easily recovered with a calibrated camera [28]. Then, under a reasonable assumption that the roll angle of the camera is $0°$ (i.e. that the camera is horizontally aligned) the IPM parameters can be recovered by using only the vanishing point of the road ahead [19], [21], [20]. Thus, we first calibrate intrinsic parameters of our camera as described in [28]. Consequently we annotate by hand the two borders of the straight road in the first image of the sequence, determine the position of the vanishing point, and calibrate extrinsic parameters. These extrinsic parameters directly correspond to the pan, tilt and height of the camera with respect to the coordinate system of the road. In order to alleviate distortions due to vehicle motion, in the future we shall design a procedure for dynamic refinement of the parameters in a way similar to [20].

A steerable filter based on a second derivative of a Gaussian is very effective in locating elongated blobs [12]. Compared to simpler extraction methods, the employed steerable filter generates not only the location of the features, but also their orientation. However, the employed filter is not scale invariant, which means that it can not be successfully applied to perspective images where the apparent width of lane boundaries changes significantly. This effect is illustrated in Figure 5.

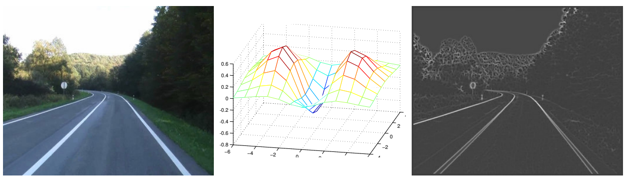
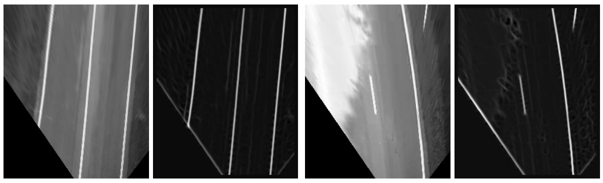


Fig. 5. Original image (left), steerable filter $G_{xx}$ (middle), and the corresponding response (right). The scale of the employed filter does not suit the lane width in the bottom of the image, resulting in a “hollow” response.

Much better results can be obtained by applying the employed steerable filter to inverse-perspective images, as shown in Figure 6(a,b). In good illumination conditions, the lanes are perfectly emphasized, and, due to their length, they should be easily estimated either as straight lines by a Hough transform, or as parabolas by a random sampling approach. Again, the estimation procedure would employ not only the positions of pixels with high response, but also the direction recovered by the steerable filter. Figure 6(c,d) suggests that acceptable results can be obtained even in the case of very uneven illumination. The right lane boundary and the middle line are nicely located, while the left lane boundary is almost invisible even to the human observer. Obviously, the steerable filter would be fooled whenever illumination effects form elongated image structures, but this did not occur so often in our experiments.



(a) (b) (c) (d)

Fig. 6. Two inverse-perspective images (a,c), and the corresponding steerable filter responses (b,d).

## V. Some results in detection of traffic signs

In this section we present experimental results in triangular traffic sign detection using a boosted cascade of Haar-like features [26], [29]. We first introduce our groundtruth acquisition method, and then present experiments with the Lienhart’s implementation of the training algorithm (cf. http://opencv.sf.net).

### A. Groundtruth acquisition

In order to train the boosted detector, one needs many images of the desired objects with annotated groundtruth positions. The strength of our project is that a large quantity of video material has already been acquired by our industrial partner. In order to exploit these resources, we developed an application which enables us to play a video, pause on frames of interest and annotate by hand the sign positions and types (cf. Figure 7). The application stores the annotated groundtruth in a separate text file and exports the corresponding frames as individual images.

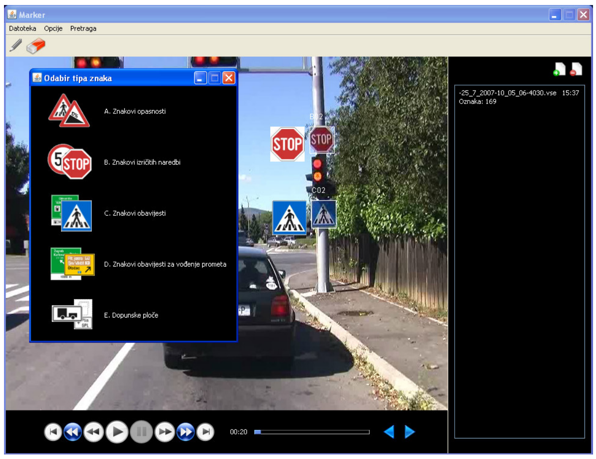



Fig. 7. A sign instance is annotated by (i) enclosing it in a bounding box, and (ii) selecting the sign type.

So far we have collected around 1000 images, corresponding to about 250 signs (our policy is to annotate each sign in several representative frames). However, this includes regulatory, warning, information, traffic guidance and other sign types, while here we are exclusively concerned with warning signs. Thus, the presented experiments are performed with a detector trained on 352 images, while the evaluation is performed on 72 images.

Some of the images from the training set are shown in Figure 8.

Fig. 8. Some of the sign images from the training set.

## B. Training

We trained four detectors which differed in the employed set of features (basic [26] and extended [29]), and in the assumption of the vertical symmetry of the detected objects (symmetric and asymmetric). The four combinations were: basic-asymmetric, extended-asymmetric, basic-symmetric, extended/symmetric. Best results were obtained by a detector which employed the original set of features [26], and assumed vertical symmetry of the detected objects. Minimal desired hit rate for each stage classifier was 0.995, while maximal false alarm rate was 0.5. We used 4000 negative examples taken from images without signs. The base resolution of the detector was 24×24 pixels.

The detectors were trained using Gentle AdaBoost [30]. Training typically took about two days on a computer with two CPUs. Both CPUs were utilized since the training program employs OpenMP (cf. http://openmp.org).

## C. Experiments and results

Best detection results were obtained by using a relatively low scale factor of the detector window (1.05). The results are summarised in Table I. The symmetric detector with the basic feature set achieved the highest detection rate of 68%, while the corresponding precision was 46%. The weak classifiers from the first five stages of the best detector are shown in Figure 9. We can easily interpret the most significant feature: it is sensitive to the bottom edge of the sign.

TABLE I
DETECTION RESULTS FOR DIFFERENT TRAINING OPTIONS.

| Detector type | Hits, misses, false alarms | | | Number of stages, classifiers | | Total time |
|---|---|---|---|---|---|---|
| basic nonsym | 44 | 28 | 40 | 16 | 120 | 25 |
| extended nonsym | 45 | 27 | 10 | 16 | 105 | 29 |
| basic sym | 49 | 23 | 58 | 17 | 148 | 16 |
| extended sym | 35 | 37 | 5 | 17 | 163 | 24 |

The detection results for individual physical signs from the test set are summarized in Table II. It turns out that most signs have been detected in at least one of images, regardless of the detector variant. This seems encouraging because in our context it might be sufficient to detect

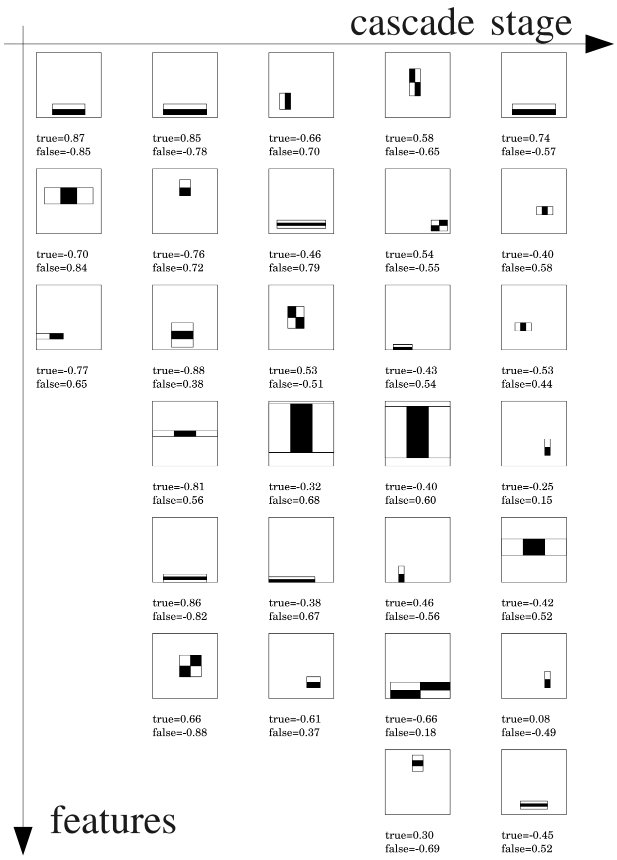



Fig. 9. The first five stages of the cascade obtained with basic set of features and symmetry on. Columns correspond to individual stages of the cascade. Numbers under each feature indicate influence of the feature to the compound result of the stage classifier, in cases of the positive (*true*) and negative (*false*) outcome. A candidate region is detected as triangular traffic sign if it passes all stage classifiers.

the sign only in one frame and use the tracking in other frames. Images of the two signs which were not detected by any detector are shown in Figure 10.

## D. A short discussion

We have presented preliminary experiments on applying the Viola-Jones object detector for the detection of traffic signs. The results are not conclusive, as we believe the results would be much better if the training set was larger. The detectors probably failed on images in Figure 10 due to the fact that the training set does not contain

TABLE II
DETECTION RESULTS FOR INDIVIDUAL PHYSICAL SIGNS.

| traffic sign instance id | number of images in test set | number of detections | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | basic set of features | | | | extended set of features | | | |
| | | non-symmetric | | symmetric | | non-symmetric | | symmetric | |
| | | 1.05 | 1.20 | 1.05 | 1.20 | 1.05 | 1.20 | 1.05 | 1.20 |
| A01#1 | 4 | 2 | 0 | 3 | 1 | 1 | 0 | 1 | 0 |
| A04#1 | 4 | 2 | 1 | 1 | 0 | 2 | 0 | 0 | 0 |
| A05#1 | 3 | 3 | 3 | 3 | 2 | 3 | 1 | 2 | 3 |
| A10#1 | 4 | 1 | 1 | 4 | 2 | 2 | 1 | 1 | 0 |
| A10#2 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 |
| A11#1 | 4 | 3 | 2 | 4 | 3 | 4 | 2 | 4 | 3 |
| A11#2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 3 |
| A11#3 | 3 | 3 | 0 | 3 | 3 | 2 | 0 | 2 | 1 |
| A11#4 | 4 | 4 | 3 | 4 | 4 | 4 | 2 | 4 | 2 |
| A12#1 | 4 | 4 | 1 | 4 | 1 | 3 | 0 | 2 | 0 |
| A13#1 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A20#1 | 4 | 4 | 2 | 4 | 3 | 4 | 2 | 4 | 1 |
| A20#2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A22#1 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A22#2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A25#1 | 5 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 1 |
| A34#1 | 4 | 4 | 3 | 4 | 3 | 4 | 4 | 4 | 3 |
| A34#2 | 4 | 2 | 1 | 4 | 1 | 4 | 1 | 2 | 2 |
| A44#1 | 3 | 3 | 1 | 2 | 2 | 3 | 1 | 0 | 1 |

enough images of such signs. Additionally, many of the training images are of poor quality due to effects of interlacing and motion blur. Nevertheless, the results seem to suggest that useful results can be obtained even with a modest training set[2].

Fig. 10. Two images in which the sign was not detected by any of the trained detectors.

## VI. Conclusions and future work

The purpose of this paper is threefold. Firstly, we present a joint research project concerned with designing computer vision modules for an existing geoinformation traffic inventory system. The two high-level vision tasks which are of greatest interest in the context of our project are the detection of surface markings and recognition of traffic signs. Secondly, we review the state of the art computer vision techniques which are relevant for solving the two tasks above. Finally, we show preliminary experimental results obtained on real images acquired from a moving vehicle.

The presented project is an example of synergy between the free enterprise and academia. We strive to create innovation by combining motivation emerging from a real context, with the specific expertise originating from previous academic achievements. It is expected that the cooperation will lead to contributions in all related fields, i.e. both academic and business, and that these contributions will result in benefits to society at large.

Preliminary results suggest that our goals are feasible, however extensive testing will be required in order to achieve robust operation on previously unseen image sequences. Further work shall be directed towards a proper enforcement of additional constraints (cf. III-C), which is expected to result in competitive recognition results.

## References


[1] L. Fletcher, N. Apostoloff, L. Petersson, and A. Zelinsky, "Vision in and out of vehicles," *IEEE Intell. Systems*, vol. 18, no. 3, pp. 12–17, 2003.

[2] M. J. Cipolloni, "Traffic control sign gis," in *Proceedings of URISA*, (Washington, D.C.), pp. 209–221, 1992.

[3] P. Arnoul, M. Viala, J. Guerin, and M. Mergy, "Traffic signs localisation for highways inventory from a video camera on board a moving collection van," in *Proc. of IV*, (Tokyo, Japan), pp. 141–146, Sept. 1996.

[4] C. Seifert, L. Paletta, A. Jeitler, E. Hödl, J.-P. Andreu, P. M. Luley, and A. Almer, "Visual object detection for mobile road sign inventory," in *Mobile HCI*, (Glasgow, UK), pp. 491–495, Sept. 2004.

[5] S. R. Madeira, L. C. Bastos, A. M. Sousa, J. F. Sobral, and L. P. Santos, "Automatic traffic signs inventory using a mobile mapping system for gis applications," in *International Conference and Exhibition on Geographic Information*, (Lisboa, Portugal), May 2005.


[2] A more complete account of our traffic sign detection experiments is going to be presented elsewhere [31].


[6] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Siegmann, H. Gomez-Moreno, and F. Acevedo-Rodriguez, "Traffic sign recognition system for inventory purposes," in *Proc. of IV*, (Eindhoven, Netherlands), pp. 590–595, June 2008.

[7] J. Borenstein, H. R. Everett, L. Feng, and D. Wehe, "Mobile robot positioning: Sensors and techniques," *Journal of Robotic Systems*, vol. 14, pp. 231–249, Apr. 1997.

[8] W. Benesova, Y. Lypetskyy, J.-P. Andreu, L. Paletta, A. Jeitler, and E. Hödl, "A mobile system for vision based road sign inventory," in *Proc. 5th International Symposium on Mobile Mapping Technology*, (Padova, Italy), May 2007.

[9] I. Dadic, "Geo-referenced video recording and analysis of road infrastructure from driver's perspective," in *Proc. WCITS*, (New York), Nov. 2008.

[10] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla, "Segmentation and recognition using structure from motion point clouds," in *Proc. of ECCV*, (Marseille, France), pp. 44–57, Oct. 2008.

[11] D. Hoiem, A. A. Efros, and M. Hebert, "Putting objects in perspective," in *Proc. of CVPR*, (New York), June 2006.

[12] J. McCall and M. Trivedi, "Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation," *IEEE Trans. ITS*, vol. 7, pp. 20–37, Mar. 2006.

[13] J. Goldbeck, B. Huertgen, S. Ernst, and L. Kelch, "Lane following combining vision and dgps," *Image and Vis. Comput.*, vol. 18, pp. 425–433, Apr. 2000.

[14] S. Šegvić, A. Remazeilles, A. Diosi, and F. Chaumette, "A scalable mapping and localization framework for robust appearance-based navigation," *Comput. Vis. Image Underst.*, vol. 113, pp. 172–187, Feb. 2009.

[15] F. Paetzold and U. Franke, "Road recognition in urban environment," *Image and Vis. Comput.*, vol. 18, pp. 377–387, 2000.

[16] C. Jung and C. Kelber, "Lane following and lane departure using a linear-parabolic model," *Image and Vis. Comput.*, vol. 23, pp. 1192–1202, Nov. 2005.

[17] A. Guiducci, "3d road reconstruction from a single view," *Comput. Vis. Image Underst.*, vol. 70, no. 2, pp. 212–226, 1998.

[18] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[19] A. Guiducci, "Camera calibration for road applications," *Comput. Vis. Image Underst.*, vol. 79, no. 2, pp. 250–266, 2000.

[20] M. Nieto, L. Salgado, F. Jaureguizar, and J. Cabrera, "Stabilization of inverse perspective mapping images based on robust vanishing point estimation," in *Proc. of IV*, (Istanbul,Turkey), pp. 315–320, June 2007.

[21] A. Catala Prat, J. Rataj, and R. Reulke, "Self-calibration system for the orientation of a vehicle camera," in *Proc. of ISPRS IEVM*, (Dresden, Germany), pp. 68–73, Sept. 2006.

[22] G. Loy and N. Barnes, "Fast shape-based road sign detection for a driver assistance system," in *Proc. of IROS*, (Sendai, Japan), pp. 70– 75, Sept. 2004.

[23] M. Garcia-Garrido, M. Sotelo, and E. Martin-Gorostiza, "Fast traffic sign detection and recognition under changing lighting conditions," in *Proc. ITSC*, (Toronto, Canada), pp. 811–816, Sept. 2006.

[24] S. Maldonado Bascon, S. Lafuente Arroyo, P. Gil Jimenez, H. Gomez Moreno, and F. Lopez Ferreras, "Road-sign detection and recognition based on support vector machines," *IEEE Trans. ITS*, vol. 8, pp. 264–278, Apr. 2007.

[25] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information," in *Proc. of IV*, (Las Vegas, Nevada), pp. 255–260, June 2005.

[26] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[27] P. Paclik, J. Novovicova, and R. Duin, "Building road-sign classifiers using a trainable similarity measure," *IEEE Trans. ITS*, vol. 7, pp. 309–321, Sept. 2006.

[28] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. PAMI*, vol. 22, pp. 1330–1334, Nov. 2000.

[29] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Proc. of ICIP*, (Rochester, New York), pp. 900–903, 2002.

[30] R. Lienhart, A. Kuranov, and V. Pisarevsky, "Empirical analysis of detection cascades of boosted classifiers for rapid object detection," in *Proc. of DAGM*, pp. 297–304, 2003.

[31] K. Brkić, A. Pinz, and S. Šegvić, "Traffic sign detection as a component of an automated traffic infrastructure inventory system," in *Proc. of AAPR/ÖAGM*, (Stainz, Austria), May 2009. accepted for presentation.