2010 13th International IEEE
Annual Conference on Intelligent Transportation Systems
Madeira Island, Portugal, September 19-22, 2010

A computer vision assisted geoinformation inventory for traffic infrastructure

S. Šegvić, K. Brkić, Z. Kalafatić, V. Stanisavljević, M. Ševrović, D. Budimir and I. Dadić

Abstract— **Geoinformation inventories are often employed as a tool for providing a comprehensive view onto the required state of traffic control infrastructure. They are especially important in road safety inspection where, in combination with georeferenced video, they enable repeatable off-line and off-site assessments as an attractive aternative to classic on-site inspection. Nevertheless, manual assessments are tedious and time-consuming even when performed off-line, and this seriously impairs the potential of the geoinformation inventory concept. This paper therefore researches a hypothesis that suitable georeferenced video processing techniques would allow reliable automation of the following operations: i) creation of the traffic inventory from the given video, and ii) assessing the video against the state in the inventory. Prominent computer vision approaches have been rigorously and systematically evaluated and the obtained results are presented. The results seem to support the hypothesis, although further work is required for a more definite answer.**

I. INTRODUCTION

Road infrastructure safety offers a large potential for significant reduction of road accidents and their consequences [1]. One of the most efficient instruments for detecting safety issues of a road network in operation is road safety inspection [2], [3]. Road safety inspection consists of periodical on-site assessments carried out by trained safety expert teams, and results in a formal report which is delivered to the relevant road authority. The most important part of the inspection concerns elements of the traffic control infrastructure: traffic signs and road surface markings [2].

The inspection procedure implies assessing current road conditions against the stored reference state. Geoinformation traffic inventories [4], [5], [6], [7], [8] are a convenient tool in achieving that, since they can provide a comprehensive insight onto the required state of a road. Traffic inventories are especially convenient when used in conjunction with georeferenced video [6] which is obtained by synchronizing image acquisition with the GPS readings as illustrated in Fig. 1. Georeferenced video allows for repeatable off-line and off-site assessments, which is more objective, more efficient and much more convenient than classic on-site inspection, especially when multiple experts are involved.

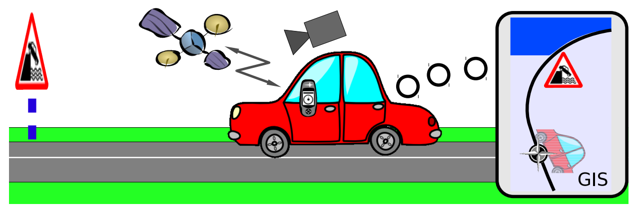

Fig. 1. Acquisition of georeferenced video for off-line road safety inspection based on a geoinformation traffic inventory. The setup enables establishing spatial relation between the individual video frames and the traffic control devices stored in the inventory.

Unfortunately, the road safety inspection is very costly in terms of expert time. This is especially true for on-site inspections, but also in the proposed off-site approach. The responsibility for the road compliance is typically assigned to local authorities, which, due to limited resources, usually delegate the implementation to external contractors [4]. The assessment frequency and the resulting safety standard therefore directly depend on the efficiency of the assessment procedure. Thus, several approaches for automating at least part of the inspection procedure have been recently addressed [6], [7], [8], [3].

In this paper, we present an existing geoinformation traffic inventory devised for road safety inspection, and propose novel computer vision techniques suitable for assessing the visibility of traffic control infrastructure. The compound software system aims at supporting automatic detection of anomalies such as broken, covered, worn-out or stolen traffic signs, and erased or incorrectly painted lanes. The presented vision components have been addressed within the frame of a research project aimed at automating i) initial creation of the inventory, and ii) assessing the actual state of the traffic control infrastructure.

The creation of a new inventory from the given georeferenced video implies: i) detecting traffic control devices and estimating their GPS locations, and ii) introducing them into the inventory (cf. Fig. 2). On the other hand, the assessment task requires that the reference state of the road is already stored into the traffic inventory, regardless of the techniques for inventory creation and maintenance. Each traffic control device from the inventory needs to be checked for presence at the corresponding location in the assessed georeferenced video (cf. Fig. 3). The creation task appears considerably less important than the assessment, since, in principle, the inventory is created only once, at the begining

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

S. Šegvić and K. Brkić and Z. Kalafatić are with the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia (name.surname@fer.hr).

V. Stanisavljević is with Tilda, Medvedgradska 1, 10000 Zagreb, Croatia (vladimir@tilda.hr).

M. Ševrović, D. Budimir and I. Dadić are with the Institute of Traffic and Communications, Kušlanova 2, 10000 Zagreb, Croatia (budimir@ipv-zg.hr, sevrovic@ipv-zg.hr).

978-1-4244-7659-6/10/$26.00

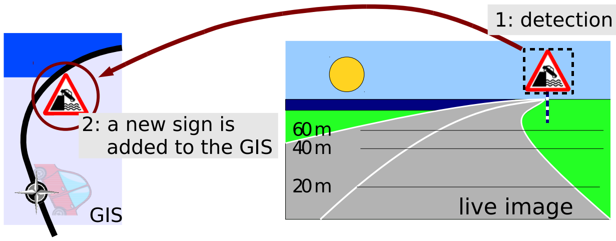

Fig. 2. Creation of a geoinformation traffic inventory: an empty inventory is populated with the detected elements of traffic infrastructure.

of the inspection service for a given road. However, in practice, being able to make a low-cost short-delay offer for a new service may prove critical for the success of the concept in a new region. The creation also may appear more demanding than the inspection, since no prior information about the road is available. However, since the creation is performed infrequently, even semi-automatic performance would be useful since the results could be checked by a human operator. On the other hand, it would make sense to assess thousands kilometers of roads on a monthly basis, which would require completely automatic behaviour, and consequently imply a much tighter error tolerance.

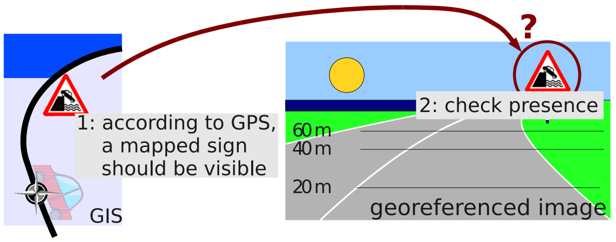

Fig. 3. Actual road state is verified against the prescribed state from the inventory. Each element from the inventory is checked for presence in recent video.

The contribution of the paper is threefold. First, we present our view onto a multidisciplinary research field of automated road inspection based on computer vision assisted geoinformation traffic inventories. Second, we provide an up-to-date review of the related work in the field. Third, we present the results of extensive, rigorous and systematic experimental evaluation of relevant techniques for detection and recognition of traffic signs. Previous accounts on this research have been previously presented to the local research communities in [9], [10].

The paper is organized as follows. Section II provides a survey of recent work in the fields of geoinformation traffic inventories and vision based detection and recognition of traffic signs. The existing geoinformation inventory developed at the partner organization is presented in Section III. Section IV describes the devised architectural solution for connecting the vision components with the traffic inventory. Section V presents experimental evaluation of the suitable computer vision techniques for detecting and recognizing traffic signs. The paper is concluded in Section VI by a short discussion of the obtained results and some directions for the future work.

II. RELATED WORK

Existing geoinformation traffic inventories differ mainly in the hardware setup used for data acquisition and in the algorithms employed for detection and recognition of traffic signs. Data acquisition is usually performed with a specially equipped vehicle (a mobile mapping system). Such a vehicle will typically have at least a camera and a differential GPS receiver. Other pieces of equipment, such as odometers, infrared cameras, inertial sensors or additional color cameras may also be employed [5], [6], [8].

Arnoul et al. [5] describe an early inventory system called AUTOCAT, which enables automatic detection and localization of traffic signs. The road is simultaneously observed with two cameras: an ordinary wide angle camera and an infrared camera. A sign is detected by combining infrared readings and color information. When a sign is detected, it is photographed using a high resolution digital camera. As GPS technology was not publicly available at the time, the 3D location of the sign is reconstructed using camera motion information provided by an inertial odometer and Kalman filtering. The photo and the 3D location are then stored to the inventory.

A system described by Madeira et al. [6] relies on videos captured by two cameras, an inertial sensor and differential GPS data to produce a georeferenced database of road signs. The signs are detected and localized using color and shape information. Absolute geographic coordinates of the identified signs are then obtained using triangulation.

Maldonado-Bascon et al. [8] describe their advances towards a fully automated inventory. Their system supports detection and recognition of a large number of sign classes. The system is able to deal with the problems of multiple detections, adverse lighting conditions and accurate localization of a sign. The detection and recognition rely exclusively on video acquired from a moving car. Experiments on a small test set of 51 traffic signs show promising detection and recognition rates.

Benesova et al. [7] suggest an alternative approach for data acquisition. Instead of using a vehicle, a pedestrian walks down the streets of interest and employs a handheld device equipped with a GPS receiver and a high resolution camera. The approach is intented to help in the cases of highly cluttered urban scenes, where it might be difficult to obtain good detection rates if the video was acquired from a vehicle.

Traffic sign detection and recognition, while essential for traffic infrastructure inventory systems, are studied in the context of many other commercial applications. One common use for them is in driver assistance systems, which are nowadays becoming common in luxury vehicles. Hence, the methods for traffic sign detection and recognition outlined below are not necessarily connected to traffic infrastructure inventory.

The majority of traffic sign detection techniques use as much prior information as possible. As the appearance of a traffic sign is strictly constrained, it is common to use color and shape cues in detector design.

Color cues can be used by performing color-based segmentation on an image. The idea is that the areas corresponding to traffic signs of target color will survive the segmentation process. The segmentation is sometimes performed in RGB color space [11], [12], [13], [14], although it often fails because RGB color space is sensitive to illumination changes. The prevalent approach is to segment the image in HSI color space [5], [6], [8], [15], [16], which should eliminate problems with illumination to some extent.

Common techniques that provide shape cues are Hough transform, fast radial transform and corner detection. Hough transform is useful in locating straight lines or circles corresponding to a sign [17]. An extension of Hough transform which generalizes it to equiangular polygons was proposed by Loy [18] for the purpose of traffic sign detection. Corner detection is sometimes performed to infer the shapes in the regions of interest [19].

Instead of using color and shape cues for traffic sign detection, one could use a general purpose detector, such as the one proposed by Viola and Jones [20]. For instance, Baro and Vitria [21] detect signs using the method of Viola and Jones augmented by a feature set proposed by Lienhart and Maydt [22]. They further filter false positives by using fast radial symmetry transform of Loy et al. [18] and principal component analysis.

Ruta et al. [23], [24] use the detector of Viola and Jones to quickly establish regions of interest in an image. They then detect equiangular polygons using an improvement of fast radial symmetry. A sign candidate is tracked using Kalman filter, a pixel relevance model and a regular shape detector.

A more detailed report on current research in sign detection is available in a recently published review paper by Nguwi and Kouzani [25].

A few machine learning techniques seem to be favored by researchers for recognition of traffic signs. These are: support vector classification [8], [16], neural networks [26], [12], cross-correlation template matching [23], [27].

Some traffic infrastructure inventory systems use tracking to enhance the results of detection and recognition. Common approaches include KLT tracker [16], Kalman filter [23], [5] or SIFT feature-based tracking [28].

To increase performance and reduce the number of false positive detections, it is useful to limit the search area of the detector to just a few regions of interest in an image. Detection of lane markings can be helpful in determining such regions. By knowing where the road is, it is easy to establish the expected positions of traffic signs. Meuter et al. [29] describe an approach to lane marking detection based on an extended and simplified fast radial symmetry transform [18] for lines. The detection is tracked using extended Kalman filters and the road is modeled by a clothoid. Nienhūser et al. [30] also model the road as a clothoid, generating its parameters by adapting a third order polynomial approximation. Their system is able to determine to which lane a specific road sign belongs to. Liu et al. [31] use Canny edge detector to detect candidate markings and then further filter them by orientation.

Compared to previous research on traffic sign detection and recognition [6], [7], [8], [23], this paper presents a rigorous and systematic study on a very large and independent evaluation set. Both training and evaluation samples have been extracted from production videos, acquired with an approximate camera speed of 60 km/h. Consequently, the resolution of our image samples is low, considerable motion blur is present, while the color is often quite unreliable. Nevertheless, we succeed in obtaining combined detection and recognition rates, in the range of over 90%, for classes in which enough training samples are available. We therefore think that this study provides a competitive insight onto the feasibility of the automatic traffic inventory concept.

III. THE GEOINFORMATION INVENTORY

Here we briefly present a geoinformation inventory for traffic infrastructure [32] which has been developed and maintained as a software product branded OptaGIS. The developed system has been a successful tool in providing commercial road inspection 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).

OptaGIS stores all information about the traffic infrastructure including traffic signs, surface markings and curbs, and presents them in the schematic view as shown in Fig. 4. The stored data is organized in several layers: aerial ortphoto, traffic signs (orange circles) with annotations (sign images), surface markings (black lines), the centerlines (yellow lines), the road identifier (orange), and the trajectory of the GPS readings (filled brown circles), and the georeferenced video. OptaGIS allows an associative look-up across different layers: the current position in the video can be set by selecting a desired GPS reading, while seeking the video pans the current view of the GIS.

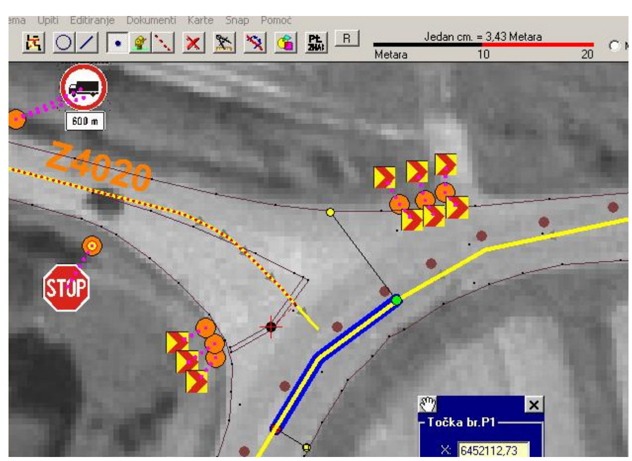

Fig. 4. The schematic view of the traffic inventory as presented by OptaGIS.

The relation between the schematic view and the georeferenced video is illustrated in Fig. 5. The live video frame shown on the right corresponds to the selected current location in the schematic view on the left. The schematic view shows the current vehicle position (red arrow), other GPS readings (green circles), mapped traffic signs (orange), and road contours (brown lines). As stated in the introduction, georeferenced video enables a deferred (off-line and off-site) and repeatable assessment, and thus offers an important alternative to the on-site inspection. The georeferenced video is acquired by a vehicle equipped with a perspective camera, 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. 5. The schematic view of the traffic inventory (left), and the corresponding georeferenced video frame (right) as presented by OptaGIS.

IV. THE DEVISED SOFTWARE ARCHITECTURE

In this section we describe the devised architectural solution for connecting the vision components with the application OptaGIS described in Section III. We first fully describe the context and the design constraints (forces) in IV-A. Then we briefly describe the auxiliary annotation tool and the development and testing framework for vision components in IV-B and IV-C respectively. Finally, we show the resulting architecture in IV-D.

A. The design constraints

OptaGIS is developed in Visual Basic in order to achieve rapid development. On the other hand, the vision components (e.g. object detection and recognition) have been developed in C++ in order to achieve maximum performance. OptaGIS therefore needs to access vision components over a dynamic library interface which we call `libmastif` in accordance with the project acronym.

It would be tedious to develop the vision components directly as parts of a dynamic library, since that would hamper testing and experimentation. We therefore opt to develop the vision components as plug-in modules (or *algorithms*) of a previously developed C++ development framework entitled `cvsh` (cf. IV-C).

It has been noted that the functionality of object detection and recognition would also be useful within the frame of the auxiliary annotation software which we call Marker (cf. IV-B). The idea is to promote an iterative approach whereby the current state-of-the-art detection and recognition algorithms are used to streamline the collection of additional training samples. However, the annotation software is written in Java and as such, unfortunately, can not simply reuse vision components over the `libmastif` interface.

B. A tool for annotating groundtruth samples of traffic signs

In order to evaluate or train algorithms for object detection and recognition, one needs many images of the desired objects with annotated groundtruth positions. One of the strengths of our project is that our industrial partner is able to provide a large quantity of road video material. In order to exploit these resources, we developed an application entitled Marker. As shown in Fig. 6, Marker enables the user to play a video, pause on frames of interest, and annotate by hand the sign positions and types. The annotated groundtruth is stored in a separate text file and exports the corresponding frames as individual images.

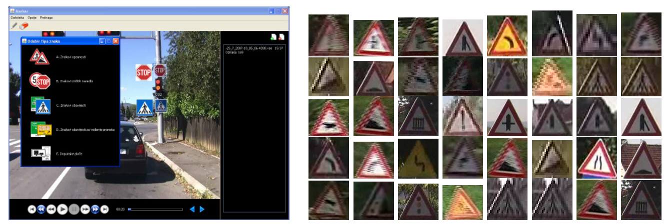

Fig. 6. A sign sample is annotated by (i) enclosing it in a bounding box, and (ii) selecting the sign type (left). Using this procedure we have manually annotated about 3000 warning signs (triangular shape, class A) (right).

It appears that best detection and recognition results can be obtained with machine learning based approaches which unfortunately require extremely large collections of training samples. In order to simplify and streamline the laborious job of groundtruth collection, Marker has been recently redesigned. The new version attempts to increase the productivity of human operators by taking advantage of previously collected data (cf. Fig. 7). The main idea of the improvement is to sustain an accelerating two-step collection procedure. Firstly, the available training collection is used to train the current generation of detection and recognition algorithms.

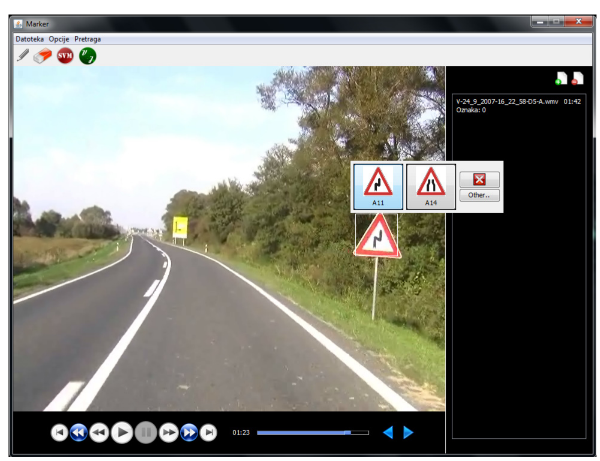

Fig. 7. After an automatic or manual detection, the new Marker performs automatic classification and offers the four most prominent classes.

Secondly, the trained algorithms are employed to perform *semi-automated* detection and classification. The processing results are verified by a human operator who accepts, rejects or corrects the proposed detection and classification. The annotated objects are consequently employed in the next iteration of the procedure, which is expected to result in a continuous performance improvement [33].

C. Development framework for vision components

The computer vision shell `cvsh` is an in-house developed C++ framework for experimenting with computer vision and image processing algorithms. The framework can be compiled both on Windows and Unix operating systems such as Linux or MacOS. It provides command-line and point-and-click user interfaces, and handles image acquisition and presentation of results. `cvsh` also offers powerful registry services which allow the application logic to be completely independent from the incorporated computer vision algorithms. Thus, users can *add* custom algorithms without having to *change* any other component in the program.

D. The proposed architectural solution

From the discussion in IV-A, it appears clear that the key architectural problem is to allow both Marker and Optagis to transparently employ `cvsh` algorithms. This can be elegantly achieved by designing the top level component of `libmastif` as a PROXY [34] towards the `cvsh` algorithms, which are simply linked with the library in their original form. In order to achieve this in a transparent fashion, `libmastif` reuses the `cvsh` registry component and thus enables hosting of an arbitrary number of `cvsh` algorithms.

Thus, OptaGIS opens a `libmastif` session by supplying the identification string of the desired algorithm. The registry component creates an algorithm instance, and afterwards the OptaGIS transparently accesses the vision components which have been independently developed and tested within `cvsh`.

In order to make the `cvsh` algorithms available to Marker, we have designed an additional dynamic library which conforms to the JNI specification [35]. In order to avoid repetition, the JNI library is designed as a simple `libmastif` proxy and is therefore named `libmastif_jni`.

The structural dependency diagram of the proposed architecture is illustrated in Fig. 8.

V. EXPERIMENTS ON TRAFFIC SIGNS

In this section we present experiments involving techniques which support the concepts of automatic GIS creation and automatic assessment of georeferenced video. We have experimented with many approaches to traffic sign detection and recognition and here we present our currently best results.

A. Methodology

The developed techniques were tested on triangular warning signs from the superclass A. The superclass A is chosen as a model subset due to its prevalence in our videos. All algorithms have been trained on the training set T2009

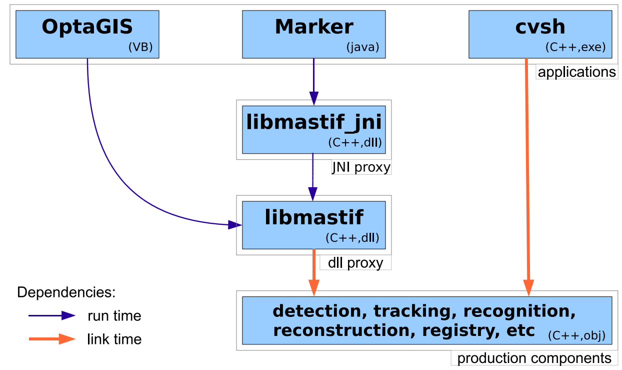

Fig. 8. The devised compound architecture for integrating independently developed vision components with i) the traffic inventory OptaGIS and ii) the annotation application Marker.

containing about 2000 triangular signs collected in 2009 using an interlaced camera. A small subset of T2009 is shown in Fig. 6 (right). The quality of the signs is rather low, while the red colour of the frame can not be distinguished as such in many samples. The algorithms have been evaluated on the evaluation set T2010, containing about 1000 triangular signs collected in 2010 using a newer progressive scan camera. The signs from T2010 have better quality, however, the colour still appears unreliable.

Both cameras feature the SDTV resolution of 720×576 and the field of view of about 48°. The viewing direction of the camera is roughly parallel to the longitudinal vehicle axis. The resolution of the signs from both sets typically range between 20×20 and 90×90 pixels.

B. Appearance-based traffic sign detection

Best detection results have been obtained with a cascade of boosted Haar classifiers [20]. Currently we employ the training tool from OpenCV [22], [36], which requires i) a large set of positive samples (cropped images of desired objects), and ii) a collection of backgrounds (images in which no objects are present). The specified number of negative samples is automatically extracted from the background images at the beginning of learning each cascade stage. Since the resulting detectors are less reliable near the basic resolution of 24×24, we perform the evaluation on subset of T2010 containing signs greater than 25×25 pixels. As shown in Table I, the obtained results heavily depend on the number of training samples N_{pos} (N_{bg} and N_{test} denote the number of background images and the number of evaluation samples). The table shows that the boosted Haar cascade detection approach achieves quite encouraging recalls when enough training samples are available.

TABLE I
IMPACT OF THE TRAINING SET TO THE DETECTION PERFORMANCE

N_{pos}	N_{bg}	N_{test}	recall	precision
352	110	72	68%	46%
898	230	918	80%	64%
2154	711	918	96%	54%

However, there are two problems which need to be addressed prior to incorporating the technique into an unsupervised technical system. The first problem is the quite high number of false positives (about one per frame). Our most recent experiments show that false positives can be reduced to a much more acceptable level of about one per 200 frames (about 9 seconds), by constraining the temporal consistency of the detector. Basically, we require that i) the object needs to be detected in many video frames, and ii) that all detections need to be similar to the first detection as judged by the Lucas-Kanade alignment algorithm [37]. It appears that a more significant removal of false positives would not be possible without a more global treatment which would distinguish the true objects by the use of context [38]. Our preliminary results along that track are presented in [39].

The second problem with the evaluated detection approach is the low localization accuracy of the detections. The problem is illustrated in Fig. 9. The presented histogram shows that the expected relative deviation of a 24×24 detection is about 1.6 pixels. We investigate the effects of this deviation to the recogintion accuracy in V-D.

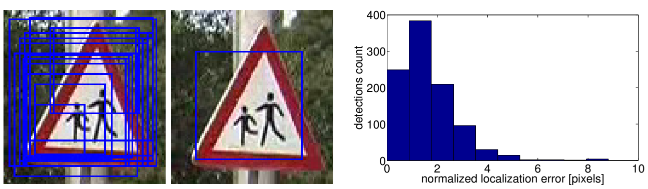

Fig. 9. The response of the grouping algorithm often deviates from the true location of the sign (left). The distribution of the relative deviations on the set T2010 normalized for the sign dimensions of 24×24 is shown on the right.

C. Colour-based traffic sign detection

Traffic signs are designed to be visible in a wide range of illumination conditions, primarily by using highly discriminative colors. Therefore, the color could be a dominant cue for traffic sign detection and many authors use it to direct the detector attention [40], [41], [42], [24].

However, the images of traffic signs captured with a video camera in a real setup show that the imaged colors vary significantly depending on lighting conditions. Some examples can be seen in Fig. 6 (right). In order to explore color properties of traffic sign images under various conditions, we conducted a statistical analysis on a representative set of images from real video. This experiment has been inspired by Jones and Rehg [43], who successfully modelled skin color by RGB histograms.

By using the annotated video sequences (Fig. 6), we collected pixel values corresponding to traffic signs of particular categories and built 3D color histograms. By using images not containing traffic signs, taken from the sequences, a background histogram was also constructed. Having histograms representing the probability distributions of the sign class and background class, a Bayes classifier [44] can be defined for segmenting individual pixels based on their color. In such a probabilistic framework, the optimal decision should be made based on the posterior probability of the object class. If we denote the measured value of the pixel to be classified as $\mathbf{x} = [\,R \quad G \quad B\,]$, the object class as ω_O and the background class as ω_B, then the pixel under consideration should be classified as object if [44]:

$$P(\omega_O|\mathbf{x}) > P(\omega_B|\mathbf{x}) \tag{1}$$

By using Bayes theorem the classification rule (1) can be expressed with class-conditional probability densities of the two classes and their priors. The pixel is to be classified as object if:

$$\frac{P(\mathbf{x}|\omega_O)}{P(\mathbf{x}|\omega_B)} > \frac{P(\omega_B)}{P(\omega_O)} = \Theta \tag{2}$$

This form of classification rule is very practical as by changing the threshold Θ the performance can be tuned. The performance of a classifier is usually expressed by its receiver operating characteristic (ROC), showing the relation between the rates of true positives and false positives as the threshold is varied.

We conducted an experiment with histogram-based segmentation of yellow panel traffic signs (route direction signs, town names), that contain mainly yellow color and black markings (arrows, text). The experiment included 3 types of histograms: (i) 3D histogram in RGB space, with bin size 1 ($256 \times 256 \times 256$ bins); (ii) RGB histogram with bin size 8 (32^3 bins); (iii) 2D histogram in normalized *rg* color space. The normalized *rgb* color space [45] is convenient because it is relatively insensitive to illumination changes. It can be used as a 2D space, because the third coordinate (usually *b*) can be computed from the other two. Yellow histograms were learned on a set of 650 annotated images, containing 2 million pixels, while the background histograms used 700 images with 287 million pixels.

The classifiers were tested on an independent set of images (200 yellow panels and 20 background images), with varying threshold Θ. The obtained ROC curves are shown in Fig. 10.

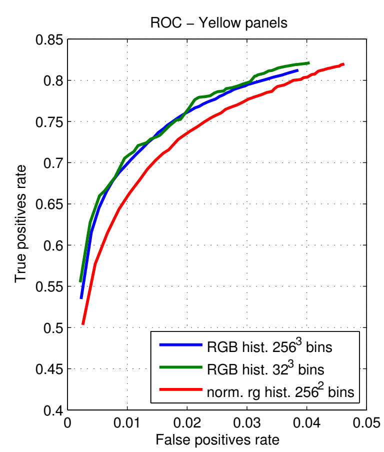

Fig. 10. ROC curves for histogram-based pixel classifier for yellow panels.

The experiment has shown that yellow color can be reliably detected with the described approach. By adding appropriate post-processing steps such as mathematical morphology and size filtering to reject small blobs of false positives and shape analysis to check the remaining candidates, a successful detector for yellow panels could be constructed.

We also tried this approach with triangular and circular signs with red rim, but with much less success. One of the reasons is probably the fact that red color occupies relatively small area of the signs. Moreover, it seems that the red pixel values in images of traffic signs very often lose their color. They become very dark, with values corresponding to brown or gray (or even violet) which are very frequently found in the background. Also, the sign images we collected from video sequences are often very small (as small as 20×20 pixels), compared to images collected by a high-resolution camera. For example, Gao et al. [41] used sign images of 400×400 pixels. However, our results suggest that for triangular and circular signs with red rim an appearance-based detector such as Viola and Jones' [20] is more appropriate than color-based approach.

D. Traffic sign recognition

The training was performed on the set T2009 with C=28 classes. The employed classificator first normalizes each training sign by i) converting it to the greyscale, ii) resizing to the common resolution of 64×64 pixels, iii) suppressing the background by a fixed triangular mask, and iv) suppressing the illumination effects by linearly stretching the histogram. The set of normalized samples is consequently projected to the (C-1)-dimensional space obtained by multiple discriminant analysis [44]. The classification is performed on the projected samples using the nearest neighbour rule [44], i.e. each unknown sample is classified as the class of the nearest projected sample from the training set [46], [47].

The evaluation is performed on a subset of the set T2010. In order to take into account localization uncertainty described in V-B, we recover the closest detection rectangle corresponding to each manually annotated sign. The annotations which were not detected have been excluded from the recognition test. We also excluded the signs from classes for which there were less than 30 training samples. At the end, the recognition evaluation set contained 906 samples in 13 classes.

The overal recognition accuracy was 91% for manual annotations and 79% for the detections. The recognition accuracy for individual classes is shown in Table II. Due to space constraints, we show only the 11 classes with most evaluation samples. The table shows that classes with more training samples (A08, A09, A10, A11, A34, A44) tend to be considerably better recognized. The effects of inaccurate detections mostly affects the classes with fewer training samples (A04, A05 and A32), leading to considerable performance hits of up to 19 percentage points.

The experiments show that large quantities of training data is required for reliable recognition. It appears that improving localization accuracy of the detections would improve recognition only for classes with less than about 50 training samples, but further research would be required to test this properly.

TABLE II

RECOGNITION OF MANUALLY ANNOTATED VS. DETECTED SIGNS.

the class appearance and id	A03	A04	A05	A08	A09	A10	A11	A17	A33	A34	A44
n(training)	41	27	31	200	223	209	242	28	35	251	212
n(testing)	128	165	156	38	48	24	47	11	206	44	32
hit(manual)	111	136	145	37	48	23	44	8	198	44	30
pct(manual)	87%	82%	93%	97%	100%	96%	94%	73%	96%	100%	94%
hit(detection)	109	104	127	37	48	23	45	10	138	43	32
pct(detection)	85%	63%	81%	97%	100%	96%	96%	91%	67%	98%	100%

VI. CONCLUSIONS AND FUTURE WORKS

This paper addressed the feasibility of applying vision-assisted geoinformation inventories for traffic infrastructure in the field of automated road inspection. Results of extensive experimental evaluation imply that, as far as triangular warning signs are concerned, at least semi-automatic detection and recognition should be possible. We believe that similar results would be achieved on any other ideogram-based class of the signs, provided that enough training samples are collected.

Although our research on indication panels is less mature than on ideogram based signs, the available results suggest that they could be detected in many instances by Bayesian color-based approaches.

Due to spatial constraits, we have refrained from presenting our experiments on detection of road surface markings. These experiments are less mature than the experiments on traffic signs, however they also suggest that automation is feasible.

None of the related approaches from the literature attempt to learn and enforce the contextual constraints which may provide important cues for improving detection and recognition. Furthermore, the presence of traffic signs is correlated with the type and the geometry of surface markings so that we believe that in the end the two detection techniques might reinforce each other. Much of our future work shall be concentrated towards exploring these research opportunities.

VII. ACKNOWLEDGMENTS

We gratefully acknowledge the contribution of Prof. Axel Pinz as a scientific advisor on this research project. We would also like to thank the reviewers for helpful comments and useful suggestions.

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] C. Joo L. C. Stefan, R. Elvik, and M. Srensen, "Road safety inspection - best practice guidelines and implementation steps," Deliverable D5 of the EU FP6 project RIPCORD - ISEREST, Tech. Rep., 2007.

[3] T. Mccarthy, N. Maynooth, C. Mcelhinney, C. Cahalane, and P. Kumar, "Initial results from european road safety inspection (eursi) mobile mapping project," in *Proc. of ISPRS CRIMT*, Newcastle, UK, June 2010.

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

[5] 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, Sept. 1996, pp. 141–146.

[6] 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.

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

[8] 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, June 2008, pp. 590–595.

[9] S. Šegvić, K. Brkić, Z. Kalafatić, V. Stanisavljević, D. Budimir, and I. Dadić, "Towards automatic assessment and mapping of traffic infrastructure by adding vision capabilities to a geoinformation inventory," in *Proc. of MIPRO*, Opatija, Hrvatska, May 2009.

[10] 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.

[11] S. Varun, S. Singh, R. S. Kunte, R. D. S. Samuel, and B. Philip, "A road traffic signal recognition system based on template matching employing tree classifier," in *Proc. of ICCIMA*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 360–365.

[12] A. Broggi, P. Cerri, P. Medici, P. Porta, and G. Ghisio, "Real time road signs recognition," in *Proc. of IV*, June 2007, pp. 981–986.

[13] W.-J. Won, M. Lee, and J.-W. Son, "Implementation of road traffic signs detection based on saliency map model," in *Proc. of IV*, June 2008, pp. 542 –547.

[14] E. Cardarelli, P. Medici, P. P. Porta, and G. Ghisio, "Road signs shapes detection based on sobel phase analysis," in *Proc. of IV*, Xi'an, China, Jun 2009, pp. 376 –381.

[15] C. Paulo and P. Correia, "Automatic detection and classification of traffic signs," in *Image Analysis for Multimedia Interactive Services, 2007. WIAMIS '07. Eighth International Workshop on*, June 2007.

[16] W. Liu, X. Chen, B. Duan, H. Dong, P. Fu, H. Yuan, and H. Zhao, "A system for road sign detection, recognition and tracking based on multi-cues hybrid," in *Proc. of IV*, June 2009, pp. 562 –567.

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

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

[19] A. de la Escalera, L. Moreno, M. Salichs, and J. Armingol, "Road traffic sign detection and classification," *IEEE Trans. IE*, vol. 44, no. 6, pp. 848–859, Dec. 1997.

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

[21] X. Baro and J. Vitria, "Fast traffic sign detection on greyscale images," *Recent Advances in Artificial Intelligence Research and Development*, pp. 131–138, October 2005.

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

[23] A. Ruta, Y. Li, and X. Liu, "Detection, tracking and recognition of traffic signs from video input," in *Proc. ITSC*, Oct. 2008, pp. 55 –60.

[24] ——, "Real-time traffic sign recognition from video by class-specific discriminative features," *Pattern Recog.*, vol. 43, no. 1, pp. 416–430, 2010.

[25] Y. Nguwi and A. Kouzani, "A study on automatic recognition of road signs," in *Proc. of CIS*, June 2006, pp. 1–6.

[26] H. Ohara, I. Nishikawa, S. Miki, and N. Yabuki, "Detection and recognition of road signs using simple layered neural networks," in *Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference on*, vol. 2, Nov. 2002, pp. 626–630 vol.2.

[27] G. Siogkas and E. Dermatas, "Detection, tracking and classification of road signs in adverse conditions," in *Proc. of MELECON*, May 2006, pp. 537–540.

[28] A. Krishnan, C. Lewis, and D. Day, "Vision system for identifying road signs using triangulation and bundle adjustment," in *Proc. ITSC*, Oct. 2009, pp. 1–6.

[29] M. Meuter, S. Muller-Schneiders, A. Mika, S. Hold, C. Nunn, and A. Kummert, "A novel approach to lane detection and tracking," in *Proc. ITSC*, Oct. 2009, pp. 1 –6.

[30] D. Nienhuser, T. Gumpp, J. Zollner, and R. Dillmann, "Recognition and attribution of variable message signs and lanes," in *Proc. of IV*, June 2008, pp. 55 –60.

[31] W. Liu, H. Zhang, B. Duan, H. Yuan, and H. Zhao, "Vision-based real-time lane marking detection and tracking," in *Proc. ITSC*, Oct. 2008, pp. 49 –54.

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

[33] S. Munder and D. Gavrila, "An experimental study on pedestrian classification," *IEEE Trans. PAMI*, vol. 28, no. 11, pp. 1863–1868, 2006.

[34] E. Gamma, R. Helm, R. Johnson, and J. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented systems*, ser. Professional Computing Series. Reading, Massachusetts, USA: Addison-Wesley, 1995.

[35] S. Liang, *Java Native Interface: Programmer's Guide and Reference*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1999.

[36] G. R. Bradski and A. Kaehler, *Learning OpenCV*. O'Reilly Media, Inc., 2008.

[37] S. Baker and I. Matthews, "Lucas-Kanade 20 years on: A unifying framework," *Int. J. Comput. Vis.*, vol. 56, no. 3, pp. 221–255, Mar. 2004.

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

[39] K. Brkić, S. Šegvić, Z. Kalafatić, I. Sikirić, and A. Pinz, "Generative modeling of spatio-temporal traffic sign trajectories," in *Proc. of UCVP*, San Francisco, California, June 2010.

[40] G. Piccioli, E. D. Micheli, P. Parodi, and M. Campani, "Robust method for road sign detection and recognition," *Image and Vision Computing*, vol. 14, no. 3, pp. 209 – 223, 1996.

[41] X. Gao, L. Podladchikova, D. Shaposhnikov, K. Hong, and N. Shevtsova, "Recognition of traffic signs based on their colour and shape features extracted using human vision models," *Journal of Visual Communication and Image Representation*, vol. 17, no. 4, pp. 675–685, 2006.

[42] Y.-Y. Nguwi and A. Z. Kouzani, "Detection and classification of road signs in natural environments," *Neural Comput. Appl.*, vol. 17, no. 3, pp. 265–289, 2008.

[43] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *International Journal of Computer Vision*, vol. 46, no. 1, pp. 81–96, 2002.

[44] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[45] L. G. Shapiro and G. C. Stockman, *Computer Vision*. Prentice Hall, 2001.

[46] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," in *Proc. of ECCV*, 1996, pp. 45–58.

[47] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. PAMI*, vol. 23, pp. 228–233, 2001.