# Traffic sign detection as a component of an automated traffic infrastructure inventory system[*]

Karla Brkić[1], Axel Pinz[2], and Siniša Šegvić[1]

[1] Faculty of Electrical Engineering and Computing
Zagreb, Croatia
*{karla.brkic,sinisa.segvic}@fer.hr*

[2] Institute of Electrical Measurement and Measurement Signal Processing
Graz University of Technology, Austria
*axel.pinz@tugraz.at*

*Abstract*

*We study the problem of traffic sign detection in the context of traffic infrastructure inventory. The data acquired during filming the roads in Croatia is presented. Based on recent approaches, and motivated by constraints present in our data, we employ the Viola-Jones object detector for triangular warning signs detection. The detector achieves correct detection rates better than 90%, which is sufficient for our application. The false positive rate is a concern, in some cases being higher than 160%, so the causes of false positive detections are analyzed in detail. We suggest a new approach of fusing the Viola-Jones detector with a priori knowledge, in form of a sign model and geometric constraints, in order to increase the correct detection rate and decrease the false positive rate.*

## 1 Introduction

Considerable research in the field of intelligent vehicles has been devoted to traffic sign detection and recognition. The majority of this research is intended to be used in driver assistance systems – systems which aim at automatically recognizing and presenting traffic information to the driver. However, our research is directed towards a different application - a system for automated traffic infrastructure inventory, similar to works [1, 17, 18, 27, 22].

Appearance of a road sign changes over time. Its colors fade, it gets occluded or damaged by bad weather conditions. Depending on the sign type, this can have moderate to severe adverse effects on traffic safety. In Croatia, local authorities of each province are therefore required to ensure that the traffic infrastructure is kept in good condition. Typically this is done by hiring external contractors that film all the roads in the province. The obtained videos are then analyzed by a human operator and every noticed malady is recorded.

Our goal is to develop a system which would automate this process. The system we propose would have two basic functions

1. automatic mapping of traffic signs by using sign detection and recognition in georeferenced video

2. verification of a newly acquired video against a previously recorded state of traffic infrastructure

The development of this system is an ongoing project carried out in cooperation with an industrial partner[1]. In this stage of the project we are primarily concerned with the problem of traffic sign detection, as it is a crucial component of the system.

This paper describes an early stage of our project and aims at exploring our data and presenting some solutions on the problem of traffic sign detection.

## 2   Related work

The vast majority of published methods for traffic sign detection take as much use of a priori information as possible. Appearance of a traffic sign is strictly constrained: the sign is always a regular polygon or a circle and its colors are well known. The exact colors used generally depend on the country, but usually include white, yellow, red, blue and black. Therefore, a typical sign detection method uses some combination of color and shape constraints to detect a sign.

Color information is usually exploited by performing color-based segmentation of the image. Such segmentation is difficult to perform in RGB space. RGB colors are very sensitive to illumination changes and traffic scenes tend to have varying illumination. Some authors [4, 9, 25, 5, 7] try to overcome this by devising simple formulas relating red, green and blue components and experimenting with appropriate thresholds. Others work in HSI [21, 12] or L*a*b [23] color spaces. There have been approaches with training neural networks [20] or support vector machines [28] for color labeling.

Shape information can be obtained by various strategies: Hough transform, fast radial transform, corner detection, pattern matching, genetic algorithms etc. Hough transform is used to locate lines or circles corresponding to a sign [12, 10]. Recently Loy and Zelinsky [16] proposed a technique similar to Hough transform called fast radial transform. It was succesfully used for sign detection in [21, 15]. Shape is sometimes determined by using corner information of candidate regions [21, 12, 7]. Some researchers use pattern matching with simple shape templates [5]. A technique based on genetic algorithms was used for detection of circular traffic signs in [24]. Shape detection often fails in cases of insufficient edge contrast, so most researchers choose to augment it with color information.

A different approach to the problem of traffic sign detection is to use a general purpose object detector instead of devising application-specific algorithms. A few researchers report success with using the robust detector of Viola and Jones [26]. For instance, Chen and Sieh [6] use the Viola-Jones detector to detect circular signs. Bahlmann et al. [2] extend the feature set described by Viola and Jones in order to use color information. Baro and Vitria [3] detect signs using the method of Viola and Jones augmented by a feature set proposed by Lienhart and Maydt [14]. They further eliminate false positives by using fast radial symmetry transform and principal component analysis. Their results are used by Escalera and Radeva [8].

The interested reader can find a more detailed report on current research in sign detection in a recently published review paper by Nguwi and Kouzani [19].

---

[1]The Institute of Transport and Communications, Zagreb, Croatia

# 3   The data

The previous section has shown that research in the area of traffic sign detection is of considerable volume. Lots of different approaches report decent results. With such an abundance of methods and sparse time, a natural question arises: how to choose which methods to try? Certainly, this depends on the exact nature of the data the method will be dealing with. For instance, there is no sense in choosing a color thresholding method if one is using a camera which provides grayscale images.

In our case, the data is 57 hours of video filmed for the purpose of road maintenance. The filming was carried out by our industrial partner, using a camera mounted on top of a car as shown in Figure 1. The data was integrated in a geoinformation system running on an on-board computer in the car, so a GPS coordinate of each frame is available.



**Figure 1: The vehicle used for acquisition of road videos. The videos are georeferenced using an on-board computer equipped with a geoinformation system.**

## 3.1   Annotating the videos

Video is not the preferable medium for testing various methods of traffic sign detection. Testing is easier when done on a collection of static images of signs. The images should be accompanied by ground truth data that describes the exact positions and types of the signs, so the evaluation of the developed methods could be done automatically. The collection should also be large enough so the obtained results bear statistical significance.

Therefore, we developed an application which enables us to annotate sign positions in different frames of a video and save annotated frames as bitmap images along with the accompanying ground truth. Annotating a sign consists of placing a tight bounding box around the sign and selecting the sign type code.

As we aim at developing a traffic infrastructure inventory system, the process of annotating has to include all traffic signs used in Croatia. Croatian regulations define five sign classes: warning signs, explicit order signs, information signs, direction signs and supplemental panels (see Figure 2).

This class division slightly differs from the known Vienna Convention on Road Signs and Signals [11] which defines eight sign categories. The class of explicit order signs contains both prohibitory signs and priority signs, as defined by the Vienna Convention. Mandatory signs, information signs and
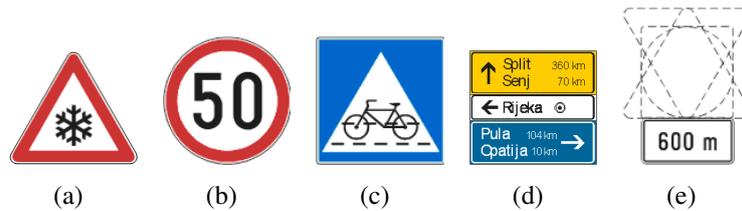
**Figure 2: Classes of traffic signs in Croatia: (a) warning sign, (b) explicit order sign, (c) information sign, (d) direction sign, (e) supplemental panel.**

special regulation signs, as defined by the Vienna Convention, are by Croatian regulations contained in a single class – information signs.

At this point, our collection contains 2352 annotated sign images. This corresponds to about 590 physical signs, as we collected four images of each sign. When a sign appears in a video, it is visible for a few seconds. The video frame rate is 24 frames per second, so there are one hundred or more frames in which the sign could be annotated. Our policy was to annotate the following four distinctive frames:

- first when the sign becomes recognizable

- last when the sign is closest to the camera

- the remaining two in between, roughly equally spaced

However, there were exceptions to this policy. In cases of very poor quality of images only one or two frames were annotated.

An example of annotating a sign in four frames in which it appears is shown in Figure 3.

## 3.2   Constraints and problems

As the videos we are using were filmed for the purpose of road maintenance, they were generally filmed when the weather was reasonably good. Of course, the task of filming for road maintenance has to be finished in reasonable time (usually one month), so it is possible for some videos to be filmed during bad weather. However, the vast majority of videos in our current collection were filmed when the weather was sunny, and all of them were filmed during daytime.

There are several factors influencing sign appearance in the videos:

**Shadows**   As the weather in our videos is generally sunny, lots of signs are partially shaded.

**Color changes**   Depending on the sun position relative to the sign, some colors may appear darkened or lightened.

**Interlacing effects**   One of the cameras with which the videos were taken uses interlacing.

**Motion blur**   Due to the motion of the filming vehicle, lots of videos suffer from motion blur.

**Figure 3: A sign annotated in four frames in which it appeared.**



(a)                   (b)                  (c)

(d)                   (e)                  (f)

**Figure 4: Examples of sign images in the database: (a) a normal sign, (b) shadow, (c) color inconsistency, (d) interlacing, (e) motion blur and interlacing, (f) occlusion.**

Figure 4 shows some examples of images from our image set.

The problems with sign appearance in the image set significantly influenced our choice of the method for sign detection. Considering unfavorable color changes frequent in our images as well as occasional appearance of shadows, we concluded that color-based methods would require some data-specific tuning to work with our collection. Effects of motion blur and interlacing might have unwanted influence on the work of shape based methods.

Our decision was therefore to first try to detect signs with a robust, general purpose detector. The possibility of using shape and color information was left for future exploration. The detector we chose was the one proposed by Viola and Jones [26].

# 4   Sign detection method and results

If two signs belong to the same semantic group, it does not necessarily mean that they are visually alike. For example, Figure 5 shows six information signs. Although all are used for conveying non-critical information to the driver, none is graphically similar to the others. Their shapes and colors vary tremendously.



**Figure 5: Examples of information traffic signs. Although their semantic use is the same – conveying information to the driver – their appearance can be very different.**

Our goal is developing a traffic infrastructure inventory system which should, by its nature, be capable of detecting and recognizing a large variety of different traffic signs. Obviously, in the final system each visually similar group of signs will have to have a dedicated detector. It will be necessary to analyze the appearance of all signs in detail and group them according to some rule depending on the detector used. For instance, if using a circular shape detector, all circular signs would be in one group, regardless of their semantic meaning.

There is, however, a semantic class of signs whose appearance is quite consistent: warning signs. Warning signs are always triangular in shape and have a thick red edge and white background. They differ only in the ideograms used. This, along with the fact that our data collection currently has more warning sign images than any other, made them ideal candidates for testing the Viola-Jones detector.

The detector was trained[2] on 824 images that contain 898 warning signs, as sometimes two warning signs appear together. The base resolution of the detector was set to $24 \times 24$ pixels, so the annotated signs were extracted from images and scaled to match that size. As mentioned previously, while annotating, a bounding box was placed tightly around the sign.[3] The training process used a set of features shown in Figure 6. We experimented with using an extended set of features proposed by Lienhart and Maydt [14], but the basic set of features proved to be better. The feature pool contained a total of 85848 features. The boosting algorithm used was Gentle AdaBoost, as experiments by Lienhart et al. [13] indicate it outperforms both Discrete AdaBoost and Real AdaBoost.
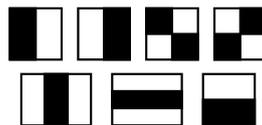


**Figure 6: Haar-like features used in training the detector.**

---

[2]The implementation from the OpenCV library was used.

[3]We performed equivalent experiments using wider bounding boxes. The hit rate remained the same, but the number of false positives doubled.

The Viola-Jones detector is a cascade of boosted classifiers. Some constraints were set on the training process for the individual stages of the cascade:

**Minimum hit rate.** Minimum hit rate for a cascade stage was set to 0.995. This means that 99.5% of all positive examples passing through the cascade stage should be detected. The training process continues to add more features to the stage classifier until minimum hit rate is reached.

**Maximum false alarm.** Maximum false alarm rate per stage was set to 0.5. Out of all detections made by a stage of the cascade, only 50% may be false positives.

**Number of stages.** 20 stages of the cascade were to be trained.

Setting the false alarm rate to 0.5 per stage and the total number of stages to 20 yields a total false alarm rate of $0.5^{20} = 9,53 \times 10^{-7}$. During training, this false alarm rate has been reached before the 20th stage, so the training was terminated. The training took three days on a server with two 3 GHz dual-core CPUs, with both CPUs being utilized. The resulting detector consists of 17 cascade stages and uses a total of 299 features.
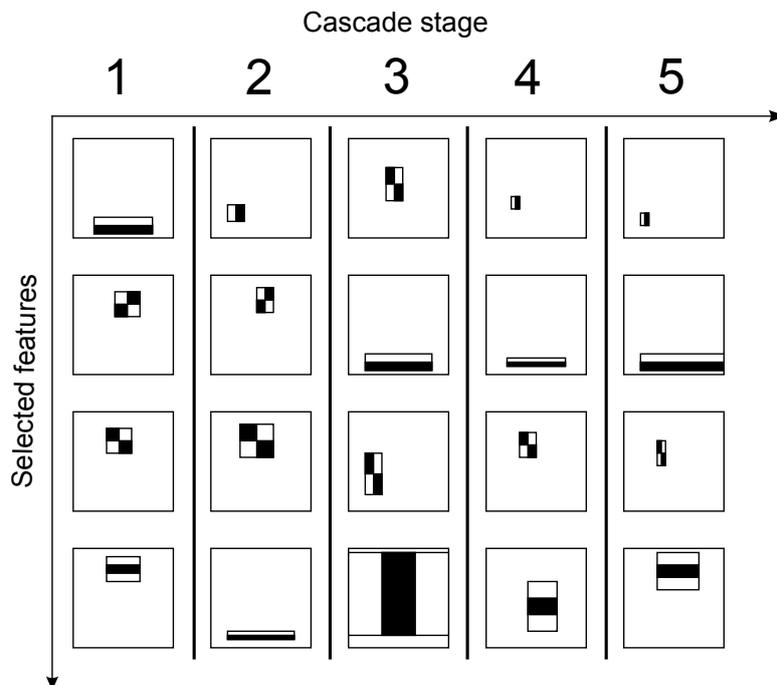


**Figure 7: Top four features of the first five stages of the Viola-Jones detector cascade. The stages are distributed on the horizontal axis, while the first four features used by each stage classifier are shown on the vertical axis.**

It is insightful to analyze main Haar-like features used by the detector. Figure 7 shows the first few features selected in training the early stages of the cascade. The first stage of the cascade uses features easily interpretable by a human: a feature sensitive to the bottom edge of a sign and three features sensitive to structure changes near the top vertex of the sign. This fact can be visualised by superimposing the features on images of typical signs, as shown in Figure 8.

The trained detector was tested on two image sets. Test set 1 consists of 91 images which contain 101 warning signs. Test set 2 consists of 68 images of 72 warning signs. None of the images from the test sets have been used in training the detector.

**Figure 8: The first four features of cascade stage 1 superimposed on sign images.**

The Viola-Jones detector works by sliding a detection window across the image. After reaching the end of the image, the window is enlarged by the scale factor and the process repeats. We tested the detector for two different scale factors: 1.05 and 1.20.

The results are summarized in Table 1. Overall, the detector performs well, achieving more than 90% hits on both testing sets. It can be seen that a smaller scale factor induces modest gains in the hit rate and significantly more false positives. Of course, using a smaller scale factor also impacts detection speed - on average, our best detector ran at 3 fps with scale factor 1.05, and at about 9 fps with scale factor 1.20.

| Test set | Scale factor | Signs | Hits [% test set] | Misses [% test set] | False positives [% test set] |
|---|---|---|---|---|---|
| 1 | 1.05 | 101 | 96 % | 4 % | 84 % |
| 1 | 1.20 | 101 | 93 % | 7 % | 42 % |
| 2 | 1.05 | 72 | 93 % | 7 % | 163 % |
| 2 | 1.20 | 72 | 90 % | 10 % | 53 % |

**Table 1: Experimental results for testing the trained detector on two different test sets. Performance was tested for detector scale factors of 1.05 and 1.20.**

The misses in detection usually ocurred when the observed signs were too far away from the camera or when only a partial detection of a sign was made. However, as stated previously, each sign was annotated in four frames, which means that for each physical sign there are four images in the database. We noticed that the signs that were missed in one image were most often detected in at least one of the remaining three. Considering that our system for traffic infrastructure inventory will work with videos, it will have the opportunity to detect a sign in more than one hundred frames. Therefore, we find the observed rate of misses in detection acceptable.

Of much greater concern is the large number of false positive detections. Succesful removal of false positives requires understanding of their origin, so false positives were analyzed in detail. It was found that the majority of false positives belong to one of the following categories: detections in trees or grass, lines on a road, triangular structures in the traffic scene, parts of traffic signs and roof structures (see Figure 9). Proportions of different false positive categories in test set 1 are shown in Figure 10.

## 5 Discussion and future work

The Viola-Jones detector shows promising results. However, the number of false positive detections is a concern. After analyzing causes for false positives, several strategies for their removal seem promising. Of course, every employed strategy will have to use the a priori knowledge of the problem.

(a)          (b)          (c)          (d)          (e)
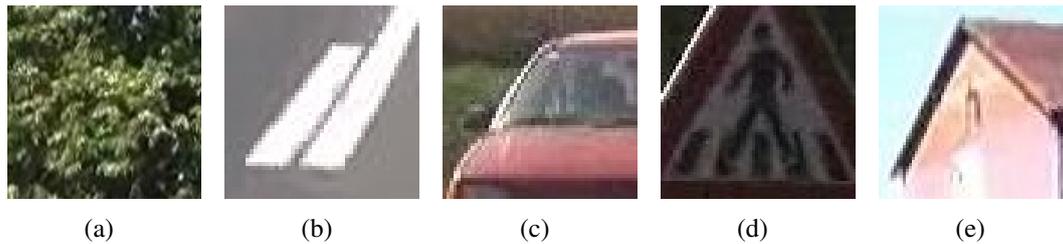
**Figure 9: False positives outlined by the detector. One image corresponds to one detected bounding box. Categories: (a) tree, (b) road line, (c) triangular structure, (d) a part of a sign, (e) roof resembling a triangle**
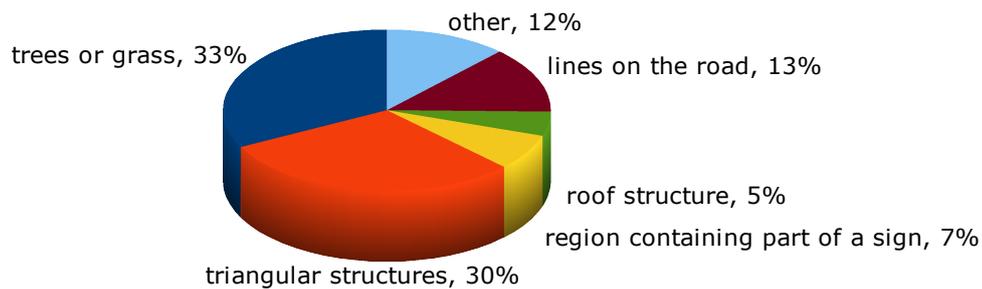


**Figure 10: Causes for false positives in test set 1, with scale factor set to 1.05**

As stated previously, appearance of a traffic sign is strictly constrained. Modeling that appearance in some way, preferably through use of shape and color, could eliminate most of the false positives. For instance, Figure 10 shows that 33% of false positives in test set 1 occur in trees or grass. In spite of undesirable color changes due to lighting, the trees and grass in our images still look predominantly green. Therefore, simply filtering predominantly green detections might be enough to eliminate 33% of false positives. Similarly, the problem of detecting lines on the road could be eliminated by filtering gray-colored detections, as the road always appears gray. This would reduce the false positive rate by another 13%. As the size of a sign is fixed, detections wider than some specific threshold could also be disregarded.

However, there are more complex cases of false positives. Triangular structures are perhaps the most difficult to eliminate. Generally, they can be positioned anywhere in the scene and have any appearance. Therefore, it is impossible to recognize them as false by using simple color constraints. We plan to eliminate them by devising a model of a sign that takes into account both shape and appearance. Attempting to fit this model would then easily detect all other simpler false positives like the aforementioned trees, as the model would certainly include color constraints. Also, using this model independently of the Viola-Jones detector and then fusing the results of the two might induce larger hit rates.

False positives could also be eliminated by introducing contextual constraints. Not only is the appearance of a sign defined very precisely, but its location in the scene is also defined very precisely. A sign can appear either by the side of the road or somewhere above the road. The size of a sign relative to the distance to the camera is also constrained: a sign that is far from the camera should appear smaller than the one close to the camera. To enforce these constraints, we first need to establish some basic geometry of the scene. The simplest way to do it is to detect the road in the image. Knowing the road

position, we could disregard detections on the road, detections too far away from the road, as well as too large detections too close to the vanishing point. Introduction of this geometry to the problem is the subject of our current research.

In studying the elimination of false positives it is necesarry to keep in mind that we are not limited to one image of a sign. The system for traffic infrastructure inventory we are developing will work with videos, which means hundreds of frames containing a single sign will be available. The system will have a tracking module which will enable tracking the detection through multiple frames. It is likely that this tracking will decrease the false detections that are results of random noise.

To conclude, we have obtained promising results in using the Viola-Jones object detector for detecting triangular traffic signs. Our further work will go in directions of adding contextual constraints and modeling sign appearance in order to reduce the number of false positives and increase the detection rate. We believe that the general purpose object detector proposed by Viola and Jones augmented with a sign model might be the solution for traffic sign detection in videos filmed in adverse illumination conditions with low-quality cameras.

# References

[1] P. Arnoul, M. Viala, J.P. Guerin, and M. Mergy. Traffic signs localisation for highways inventory from a video camera on board a moving collection van. *Intelligent Vehicles Symposium, 1996., Proceedings of the 1996 IEEE*, pages 141–146, Sep 1996.

[2] C. Bahlmann, Y. Zhu, Visvanathan Ramesh, M. Pellkofer, and T. Koehler. A system for traffic sign detection, tracking, and recognition using color, shape, and motion information. *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*, pages 255–260, June 2005.

[3] X. Baro and J Vitria. Fast traffic sign detection on greyscale images. *Recent Advances in Artificial Intelligence Research and Development*, pages 69–76, October 2004.

[4] M. Benallal and J. Meunier. Real-time color segmentation of road signs. *Electrical and Computer Engineering, 2003. IEEE CCECE 2003. Canadian Conference on*, 3:1823–1826 vol.3, May 2003.

[5] A. Broggi, P. Cerri, P. Medici, P.P. Porta, and G. Ghisio. Real time road signs recognition. *Intelligent Vehicles Symposium, 2007 IEEE*, pages 981–986, June 2007.

[6] Sin-Yu Chen and Jun-Wei Hsieh. Boosted road sign detection and recognition. *Machine Learning and Cybernetics, 2008 International Conference on*, 7:3823–3826, July 2008.

[7] A. de la Escalera, L.E. Moreno, M.A. Salichs, and J.M. Armingol. Road traffic sign detection and classification. *Industrial Electronics, IEEE Transactions on*, 44(6):848–859, Dec 1997.

[8] S. Escalera and P. Radeva. Fast greyscale road sign model matching and recognition. *Recent Advances in Artificial Intelligence Research and Development*, pages 69–76, 2004.

[9] L. Estevez and N. Kehtarnavaz. A real-time histographic approach to road sign recognition. *Image Analysis and Interpretation, 1996., Proceedings of the IEEE Southwest Symposium on*, pages 95–100, Apr 1996.

[10] M.A. Garcia-Garrido, M.A. Sotelo, and E. Martm-Gorostiza. Fast traffic sign detection and recognition under changing lighting conditions. *Intelligent Transportation Systems Conference, 2006. ITSC '06. IEEE*, pages 811–816, Sept. 2006.

[11] Inland transport comitee. *Convention on road signs and signals.* Economic comission for Europe, 1968.

[12] Wen-Jia Kuo and Chien-Chung Lin. Two-stage road sign detection and recognition. *Multimedia and Expo, 2007 IEEE International Conference on*, pages 1427–1430, July 2007.

[13] Rainer Lienhart, Alexander Kuranov, and Vadim Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In *In DAGM 25th Pattern Recognition Symposium*, pages 297–304, 2003.

[14] Rainer Lienhart and Jochen Maydt. An extended set of Haar-like features for rapid object detection. In *IEEE ICIP 2002*, pages 900–903, 2002.

[15] Gareth Loy. Fast shape-based road sign detection for a driver assistance system. In *In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS*, pages 70–75, 2004.

[16] Gareth Loy and Alexander Zelinsky. A fast radial symmetry transform for detecting points of interest. In *In: 7 th Euproean Conference on Computer Vision*, page 358. Springer, 2002.

[17] S. Madeira, L. Bastos, A. Sousa, J. Sobral, and L. Santos. Automatic traffic signs inventory using a mobile mapping system for GIS applications. In *International Conference and Exhibition on Geographic Information*.

[18] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Siegmann, H. Gomez-Moreno, and F.J. Acevedo-Rodriguez. Traffic sign recognition system for inventory purposes. *Intelligent Vehicles Symposium, 2008 IEEE*, pages 590–595, June 2008.

[19] Y.Y. Nguwi and A.Z. Kouzani. A study on automatic recognition of road signs. *Cybernetics and Intelligent Systems, 2006 IEEE Conference on*, pages 1–6, June 2006.

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

[21] C.F. Paulo and P.L. Correia. Automatic detection and classification of traffic signs. *Image Analysis for Multimedia Interactive Services, 2007. WIAMIS '07. Eighth International Workshop on*, pages 11–11, June 2007.

[22] Christin Seifert, Lucas Paletta, Andreas Jeitler, Evelyn Hödl, Jean-Philippe Andreu, Patrick Morris Luley, and Alexander Almer. Visual object detection for mobile road sign inventory. In *Mobile HCI*, pages 491–495, Glasgow, UK, September 2004.

[23] G.K. Siogkas and E.S. Dermatas. Detection, tracking and classification of road signs in adverse conditions. *Electrotechnical Conference, 2006. MELECON 2006. IEEE Mediterranean*, pages 537–540, May 2006.

[24] A. Soetedjo and K. Yamada. Fast and robust traffic sign detection. *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, 2:1341–1346, Oct. 2005.

[25] S. Varun, Surendra Singh, R. Sanjeev Kunte, R. D. Sudhaker Samuel, and Bindu Philip. A road traffic signal recognition system based on template matching employing tree classifier. In *ICCIMA '07: Proceedings of the International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007)*, pages 360–365, Washington, DC, USA, 2007. IEEE Computer Society.

[26] Paul Viola and Michael Jones. Robust real-time object detection. In *International Journal of Computer Vision*, 2001.

[27] J.Ph. Andreu W. Benesova, Y. Lypetskyy, 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 2004.

[28] Shuangdong Zhu and Lanlan Liu. Traffic sign recognition based on color standardization. *Information Acquisition, 2006 IEEE International Conference on*, pages 951–955, Aug. 2006.