Proceedings of the
2007 IEEE Intelligent Vehicles Symposium
Istanbul, Turkey, June 13-15, 2007

Road Sign Detection from Edge Orientation Histograms

Bram Alefs, Guy Eschemann, Herbert Ramoser, Csaba Beleznai

***Abstract*—This paper presents a system for road sign detection based on edge orientation histograms. Edge orientation histograms are reliable, scale and contrast invariant features that can be extracted efficiently using integral images. A learning method is introduced that selects features based on the implicit transmission function of the designer's template to the object's appearance in the image. The system is able to detect 85% of the objects on from 12 pixels width and 95% for objects on from 24 pixels width at a low false alarm rate.**

I. Introduction

ROAD sign detection is mainly interesting for driver assistance systems, in case the human detection performance can be reached. The main reason for this is that the size and duration of visibility for normal driving conditions are thus far minimized that even human may miss signs if he/she is not fully attending. This paper discusses a monocular vision system for reliable road sign detection at a range that is sufficiently large to perform interactions, based on a method for feature selection and matching using edge orientation histograms. The desire for detection and classification of road signs is relatively old. However, current image processing methods still fail to solve the main problem given the underlying issues for road sign detection. These issues include (1) poor quality of image data, especially of color at large distances for conventional camera systems, (2) a fast detection procedure that determines the true object out of many potential object positions as required for real world applications and (3) design and traffic related dynamics that are highly optimized for (excellent) human performances on computer vision. This paper aims to solve two out of three issues by using a method for fast matching of edge orientation histograms.

First, a distinction is made between detection and classification. Detection tells where in the field of view a road sign is situated. Typically, detection uses features that are specific to the road sign itself and justifiably it ignores contextual information. Classification indicates which type of sign is present, at any desired level of detail, including optical character recognition for reading numbers and text. In fact, most features that are used for classification are used also for detection, and classification can be seen as a sub task of the detection process. This paper discusses features related to detection of the different road signs types, which also may provide a good basis for classification.

Second, road signs are signals, on purpose put there by man, with the desire to be optimally well visible. This results in the typical geometric shape and a highly reflective and homogeneous colored planar surface. The human vision is able to detect such object at far distances and large clutter of other (more natural) objects. For computer vision, detecting such a precisely defined object, is a rather specific task, and different from more general tasks, such as detection of all types of vehicles with a large variety in shape and appearance. In case of the vehicle, basic properties need to be extracted from the appearance, e.g. by evaluating a large set of sample images, or by human intuition. In case of road sign detection the original template (signal) is known, and in fact, machine learning methods can be applied to learn the possible transmission of this signal. This approach is different from general detection method, as used for vehicles, where all objects invariants need to be extracted from the image in an unsupervised way. Using prior knowledge, this paper proposes a method for learning the implicit transmission functions of a priori known object template:

$$T_{\{X,Y,Z,R,..\}} : S_{Type} \rightarrow I(x,y) \qquad (1)$$

where, $T_{\{X,Y,Z,R,..\}}$ is the transmission function with specific context dependent parameters, camera position, rotation, illumination, occlusion and manufacturing conditions. The function is denoted as implicit, since in praxis, the explicit parameters cannot be retrieved from the image data. S_{Type} is the explicit road sign model as defined by designer's template and $I(x,y)$ is the road sign as it appears in the image with horizontal and vertical coordinates, x, y respectively.

Third, the human eye performs excellent on detecting noisy objects and rules out any camera system in the trade-off between a large field of view and a high spatial resolution. Due to the low resolution of camera systems, reliable features for road sign detection *cannot* be based on

Manuscript received Januari 12, 2007.

Bram Alefs is with Austrian Research Centers GmbH-ARC. Donaucitystrasse 1, 1220 Vienna, Austria (bram.alefs@arcs.ac.at)

Guy Eschemann, Herbert Ramoser and Csaba Beleznai are with Advanced Computer Vision ACV-GmbH. Donaucitystrasse 1, 1220 Vienna, Austria. www.smart-systems.at

detailed modeling of the object's shape. Instead, state-of-the-art object descriptors include statistically relevant features, starting from the single Gaussian parameters such as mean, variance to local shape descriptors, such as edge orientation histograms and more complex SIFT. This paper proposes a method for matching edge orientation histograms (EOH) for local object regions. Edge orientation histograms can be calculated efficiently using integral images and, moreover, they are robust for small variations in position and rotation. It emphasizes the use of a constitution of local indicators in a similar fashion as weak classifiers for AdaBoost related detection methods.

The paper is organized as follows. Section II discusses previous work on road sign detection. Section III discusses the method for feature extraction, learning and evaluation strategies in detail. Section IV presents the system layout using a color camera and user interfaces. Section V presents results on real world driving scenarios within city.

II. Related work

Proposed methods on road sign detection concentrate on color [1-4], local and global shape features [5-8] and several learning methods [8-11], including extensions to text recognition [12,13]. Learning methods based on local invariant statistical descriptors include color variation [14] and wavelets [15]. The latter uses responses of wavelets, using Haar-like features shapes and a boosted cascade [16-18]. The main advantages of such a classifier are the fast feature evaluation using integral images and real-time performance using a decision tree. Most interesting of [15] is the joint evaluation for features of different road sign types, including both, general and type specific features in one equation.

$$y_i = sign\left(\sum_{t=1}^{T} \alpha_t sign\left(\langle f_t, x_i \rangle - \theta_t \right) \right) \quad (2)$$

Where, $y_i > 0$ indicates a positive object, and $\langle f_t, x_i \rangle$ indicates the distance measure between feature vector f_t and observation vector x_i, including shape and color entries. α_t and θ_t are derived by a boosting procedure. Two remarks have to be made. First, multi-class AdaBoost has been topic of investigation recently [19-21] and knowledge of individual modalities of the training set a priori may help prevent cyclic behavior between features of concurrent modalities [22].

Second, as for other learning based methods, bottom up learning of appearances require a large variety in the training set. In terms of equation (1), variety of $T_{\{X,Y,Z,R,..\}}$ is difficult to determine only out of the sample space $I(x,y)$, even if many images are used. This paper discusses an alternative, using semi-supervised learning that starts first from the designer's template S_{Type}. Second, the method uses training samples to select features that are invariant of the transmission function $T_{\{X,Y,Z,R,..\}}$.

Fig. 1. Designer's templates for five types of road signs (luminance only, from left to right): Give Way, Priority Road, Stop, No Entry, Speed Limitation (the number is mirrored in order to mimic typical number like structure).

Fig. 2. Division in 5, 4, 4, 5 and 4 sub regions, respectively, for edge orientation histogram determination.

III. Road Sign Learning and Detection

The proposed system detects 5 road sign types in parallel. Currently detected signs include: Give Way, Priority Road, Stop, No Entry and Speed Limitation. Fig. 1 shows the templates as obtained from road legislation for the local area. The template for speed limitation is mirrored, indicating an arbitrary number in this area. The learning procedure is done in two steps.

A. Training

First, candidate edge orientation histograms are extracted for each of the designer's templates, in the same way as done for evaluation (see below). Fig. 2 exemplifies different characteristic regions for a total of 5, 4, 4, 5 and 4 sub regions, for the different road signs, respectively. For reasons of fast evalution, the regions are extended to 3x3 blocks on a regular grid basis. Each block covers half of the object width and height (as shown by the 2nd, 3rd and 5th template of fig. 2), but includes also intermediate positions, with half block-size overlap. Each block results in one template vector f_t, consisting of the edge orientation histogram for all the template's pixels within the block. The set of templates are used for matching to the observation vector x_i, in the image. The combination of neighboring image regions result in the observation value for each position in the image:

$$C(x,y) = \prod_i (C_i)^{\alpha_i} \quad (3)$$

Where $C(x,y)$ is the matching result at position {x,y}, α_i are the set of weights for each of the templates, $0 \le \alpha_i \le 1$ and $C_i = \langle f_t, x_i \rangle$ is the dot product between template vector f_t and the observed edge orientation histogram x_i for the local image region. Note that f_t and x_i are properly normalized and $0 \le C_i \le 1$. Since α_i and C_i are normalized, $C(x,y)$ is, and $C(x,y) \ge D$, where $0 \le D \le 1$ is the detection threshold. Now, the set of

weights $\{\alpha_1, \alpha_2, \ldots \alpha_n\}$ for all n sub regions, can be determined based on the transmission for each of the template vectors individually, and $C(x,y) = \left(\prod \langle f_t, x_i \rangle\right)^{1/n}$ in case $\alpha_i = 1/n$ for all of totally of n selected features.

Second, the values α_i, comprising the transmission function $T_{\{X,Y,Z,R,..\}}$ are trained using a validation set of example images. Fig. 3 shows typical appearance of different road sign types after transmission. The upper row shows examples for the type priority road, including rotations (3rd) and low contrast differences (4th and 5th). The second row shows various types, including inhomogeneous regions (2nd and 5th) , and low contrast (4th). The third row shows examples of speed limitation, including bleached color (1st), motion blur (3rd) and over saturation (5th).

Images at different instances of totally 32 road signs were used to validate the transmission of the template vectors. Training is performed by evaluation 3x3 sub regions as discussed above, with vectors consisting of edge orientation histograms with 8bins. The resulting classification space is restricted to 3x3x8=72 dimensions, totally. Increase of the vector length, e.g. by a higher region density is avoided. This has two reasons: (1) the minimal amount of pixels per region would require a better object resolution (which is not available at large distances) and (2) the evaluation complexity increases. Extension of the vector length by adding additional object features, like color, did not lead to substantially better results, especially in cases where the object was small in the image and the apparent red rims of the Give Way and Speed Limitation signs where strongly eroded by neighboring pixels.

In order to determine the values α_i, a set of validation images for each sign type at all relevant scales was designed. For each scale, a match to each feature vector was made and the median response was estimated. As reference, the response to a set of randomly chosen image regions, containing no road sign, were used. Because of the relative low dimensionality of the classification space, α_i were set in order to determine a binary decision: α_i=1 in case $R_{Pos} \geq 1.5 R_{Neg}$, where R_{Pos} is the median response on the positive training samples and R_{Neg} is the median response on the reference set, and α_i=0 otherwise.

B. Evaluation

Given the set of trained classifiers, the input image is scanned for presence of each one of the road signs individually. This is done in three steps. First, starting from a gray coded image, a Gaussian pyramid is built and for each layer a region of interest is selected. The region of interest is selected so that full resolution is only evaluated

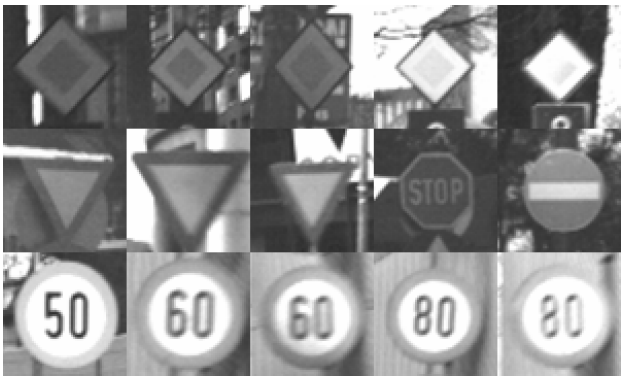

Fig. 3. Typical appearance of road signs, recorded from a driving vehicle (120x120 pixels).

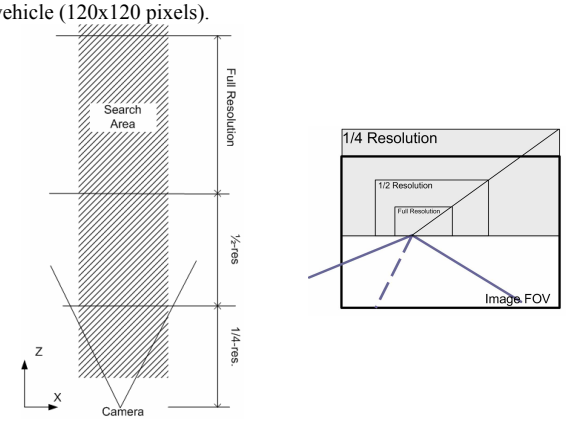

Fig. 4. Search space reduction. Left: top view with rectangular search region. Right: projected of band regions with different resolutions on the image

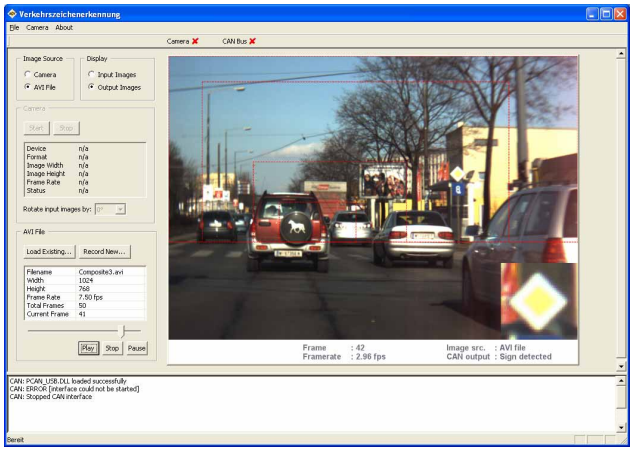

Fig. 5. Graphical User Interface

near to the horizon, whereas uppermost pyramid levels are evaluated for the entire field of view. Fig. 4 indicates this multi-resolution approach graphically. Left, it shows the search area from a top-view perspective. Right, it shows the ROI for different pyramid levels. For the lower two pyramid levels (higher resolution) only part of the field of view is evaluated. The search area is chosen so that all road signs on the right side of the vehicle are detected, even incurves and multilane scenarios with changes in the pitch angles due to acceleration and braking.

For all specified search regions, the gradient strength and orientation is determined using central differential filters, [-1

0 1] and [-1 0 1]^T, in order to determine horizontal and vertical image gradients, respectively. For each position, the gradient orientation is determined as the angle between the horizontal and vertical gradient strength. The edge orientation is determined by quantification of the gradient orientation to 8 bins that cover the contrast invariant range of edge orientations between 0 and 180deg. For each bin, an integral image is determined by updating the following equation:

$$I_k(x,y) = G_k(x,y) + I_k(x-1,y) + I_k(x,y-1) - I_k(x-1,y-1) \quad (4)$$

Where I_k is the integral image for orientation k, $I_k(x,0) = I_k(0,y) = 0$ and $G_k = |G_x| + |G_y|$ is the local gradient strength, in case the pixel has orientation k, $G_k = 0$ otherwise. Now, for a given region with coordinates {x, y, w, h}, the edge orientation histograms can be extracted by evaluating:

$$H_k(x,y,w,h) = I_k(x+w,y+h) - I_k(x+w,y) - I_k(x,y+h) + I_k(x,y) \quad (5)$$

For each integral image I_k, and normalizing the output vector so that $\sum_k H_k^2 = 1$.

IV. System Layout

The system is implemented as a graphical application running under Microsoft Windows XP. It was developed using Microsoft Visual C++ and the Microsoft Foundation Class (MFC) library. The system reads input images from either a color camera through an IEEE1394 connection, or a pre-recorded Audio Video Interleave (AVI) file. Information about the detected road signs are displayed on the screen, and sent to the vehicle's Controller Area Network (CAN) through a USB to CAN interface.

The Graphical User Interface (GUI) continually displays the input camera image along with the position of the detected road signs as overlaying bounding boxes. It allows for easy configuration of the system through menus and dialog boxes. Additionally, the system provides a few (hidden) engineering features, such as displaying the position of the road sign hypotheses, or recording the input (unprocessed) and/or output (with bounding boxes) image streams to AVI files.

The underlying application consists of three threads running in parallel: image acquisition, image processing and CAN interface. The image acquisition thread manages the camera interface and feeds the image processing thread with new images. The image processing thread, which performs the actual road sign detection on the images provided by the image acquisition thread, makes heavy use of the Intel Integrated Performance Primitives (IPP) for optimal performance. The CAN interface manages the communication with the vehicle network according to a proprietary transport protocol. The system achieves real time performance at 7.5 frames per second on a Laptop PC including post-processing, online visualization of the results and a search range for all objects between 20 and 160pixels width.

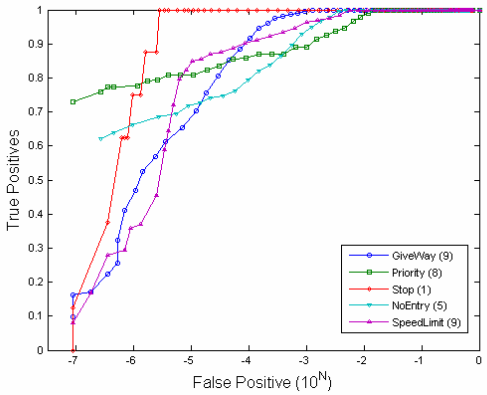

Fig. 5. Receiver operation characteristics (ROC) for typical road sign examples.

V. Results

Results are obtained in two steps, for driving scenarios around the city of Vienna, Austria. First, results were evaluated off-line, for sequences of a rather high quality camera. Second, on-line results were evaluated, for a camera system that meets current automotive constraints. For off-line evaluation, a color camera is mounted behind the windshield of a test vehicle with a field of view of about 35deg horizontally. Images were captured at a size of 1024x768 pixels, horizontally and vertically, respectively. Images were converted to gray levels and a Gaussian pyramid was created. Due to noise of the camera, color coding etc., images showed sufficient quality on from the second pyramid level. Regions of interest were defined comprising maximally 256x80 pixels in horizontal and vertical dimension, respectively. For this configuration, the third pyramid level (1/4 resolution) already covers the entire field of view in horizontal direction (see also fig. 4). For each region of interest, 8 integral images were determined, comprising different edge orientations. Feature vectors were extracted and matched to the templates according to equation (3). Fig. 5 shows the receiver operation characteristics for the average of the different object types in typically cluttered background. The different lines show that a false alarm rate of less than 10^{-4} or 10^{-3} (0.01-0.1%) can achieved at 80-90% detection rate.

For evaluation of each pyramid level, sub regions are extracted at 6x6 and 8x8 pixels in width and height, using half size overlap. These settings results in a total of 1844 and 960 image positions evaluated for each level and both scales, respectively. Without any further processing, the detection includes one or two false alarms per frame, depending on the amount of background clutter. Two post-processing methods are used in a cascaded way, using

features additional to the edge orientation histograms. First, spatial and temporal smoothing of the detection map rules out the majority of "sporadic" false detections. Second, remaining detections were verified for symmetry in positive Cr response (red, all signs) and negative Cb (yellow, Priority Road only) response of the YCbCr color space extracted for the down sampled images. Although, the false alarms from the detection show similar EOHs, they were not consistent over space-time and did not show proper color symmetry. After post-processing false alarms were only found sporadically, at locations of dense background clutter, such as advertisement panels.

For online evaluation, a camera type was chosen at VGA-resolution (640x480pixels) and a small-size lens mount (12mm-type) with 8mm focal length. Due to the lens mount, images contained much blur, and detection was only applied on from the second pyramid level (320x240 pixels, width x height). The system was evaluated online in urban areas at different driving occasions, including different illumination conditions, clutter and occlusion. Totally about 8h of driving were required to obtain recordings for a set of 105 relevant road signs. Detection results were stored along with the image data for off-line statistical evaluations. For each sign, the sequence of the detected image patch was stored. The sequence several appearances as follows from change in size, position and rotation as follows from the range and the vehicle pitch, roll and yaw. Signs are collected for following shapes: circular (speed limitations, no entry, forbiddance), triangular (give way), diamond (priority road) and octagonal (stop sign). Depending on the driver's speed, the number of relevant frames ranged between 10 and 200 per sign (recorded at 7.5fps).

Totally 16 of the 105 road signs were missed. This was mainly due to 1) insufficient visibility, 2) rotations and 3) clutter. Insufficient visibility was mostly caused by over saturation of the road sign region and lack of temporal presence in case of a fast turn. The temporal integration was set, so that road signs were required to be visible at least for three subsequent frames. Rotations and deformations were found for many occasions, and only a part of the rotated signs was missed. Clutter and partial occlusions were mainly caused by stickers and elongated objects like poles. Some sporadic false alarms occurred at specific occasions, mainly in strongly cluttered background of inner city scenarios.

The remaining 89 detected signs include red-rimmed circular shapes other than speed limitations. These can be eliminated using simple classification methods. However, in order to show generality of the detection method, these are included for the statistical evaluation. Totally, 11 signs for speed limitations were detected and 35 for other forbiddance, of which 18 for no parking. The other target signs include 16 for give way, 8 for priority road, 7 for stop and 12 for no entry. For each road sign, the width in the image at first detection was determined with an accuracy of 2 pixels and the 30% and 70% point of the distribution was derived. Table 1 shows the results.

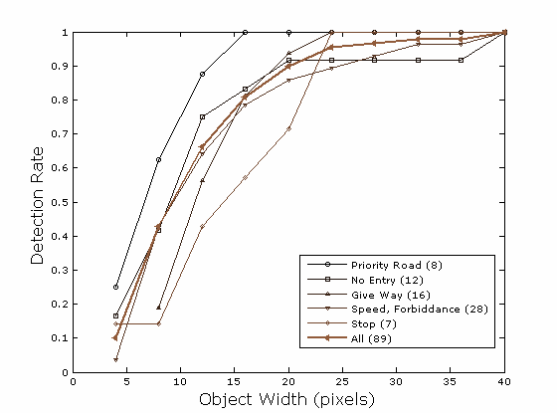

Fig. 6. Object width at first detection, for different road sign types. Widths relates to image at second pyramid level (320x240 pixels).

Object Width (Image: 320x240pixels)	#Signs	Width 30% (pixel)	Width 70% (pixel)
Priority Road	8	8	12
No Entry	12	8	14
Speed Limit*	28	10	16
Give Way	16	12	16
Stop	7	14	20
All	*71*	*10.3*	*15.6*

Table 1. * including forbiddance other than no parking

Depending on illumination conditions, signs were detected on from 8 pixels width, only few were first detected if larger than 24 pixels width. For totally 89 signs (including no parking) 30-70% is detected between 10 and 16 pixels width and 95% is detected on from 24 pixels width. Fig. 6 shows these results graphically. Fig. 7 includes an enlargement of the detected area in the lower right corner. Please note the inhomogeneous surface for the different regions, if enlarged on the electronic version of this paper.

VI. Conclusions

This paper present a method for road sign detection based on weighted matching of edge orientation histograms. The method aims to detect the road sign as early as possible, using a camera system as available for automotive industry. A classifier was designed that combines weak indicators for sub regions using a learning procedure for the implicit transmission function of the designed template to the image appearance. Results show 85% detection rate of objects on from 12 pixels width and 95% on from 24 pixels, for largely cluttered scenes in real time driving scenarios. False alarms were reduced by post-processing, to strongly cluttered scenes, and run-time can largely be improved by reducing the search range (currently 7.5fps for full coverage between 20 and 160pixels object width).

Acknowldegment

The authors thank Markus Clabian and Maike Lohndorf for their useful input and comments.

Fig. 7. Results for different road sign types at detection. The insert on the lower right corner shows an enlargement show of the object.

REFERENCES

[1] X.W. Gao et al. "Recognition of traffic signs based on their color and shape features extracted using human vision models" Journal of Visual Communication and Image Representation, 2006.

[2] V. Moreno, A. Ledezma, A. Sanchis "A static image based-system for traffic sign detection"", proc. int. multi-conf. Artificial Intel. and applications IASTED, 2006.

[3] A de la Escalera, J.M. Armingol, M. Mata: "Traffic sign recognition and analysis for intelligent vehicles" Image and Vision Computing, 2003

[4] J. Torresen. J.W.Bakke, L.Sekanina, "Efficient recognition of speed limit signs" proc. IEEE Intel. Vehicle Symposium, 2004.

[5] A. de la Escalera et al. "Road traffic sign detection and classification" IEEE trans. on intel. transportation systems, 1997.

[6] M. A. Garcia-Garrido, M. A. Sotelo, E. Martin-Gorostiza, "Fast traffic sign detection and recognition under changing lighting conditions" proc. IEEE intel. vehicle symposium, 2006.

[7] N. Barnes, A. Zelinsky "Real-time radial symmetry for speed sign detection" proc. IEEE Intel. Vehicle Symposium 2004.

[8] P. Gil-Jimenez et al, "Traffic sign shape classification evaluation II: FFT applied to the signature of blobs" proc. IEEE Intel. Vehicle Symposium, 2005.

[9] P. Paclik, J. Novovicova, R. Duin "Building road-sign classifiers using trainable similarity measure" IEEE trans. on Intel. Transportation Systems, 2006.

[10] A. de la Escalera et al. "Visualsign information extraction and identification by deformable models for intelligent vehicles", IEEE trans. on Intel. Transportation Systems, 2004.

[11] S. Lafuenta-Arroyo et al, "Traffic sign shape classification evaluation I: SVM using distance borders" proc. IEEE Intel. Vehicle Symposium, 2005.

[12] Y. Liu, T. Ikenaga, S. Goto, "Geometrical, physical and text/symbol analysis based approach to traffic sign detection system" proc. IEEE Intel. Vehicle Symposium, 2006.

[13] W. Wu, X. Chen, J. Yang, "Detection of text on road signs from video" IEEE trans. Intel. Transportation Systems, Vol6./4, 2005.

[14] T.T. Zin, H. Hama, "Robust road sign recognition using standard deviation" proc. IEEE conf on Intel Transportation Systems, 2004.

[15] C. Bahlmann et al. "A system for traffic sign detection tracking and motion information" proc. IEEE Intel. Vehicle Symposium, 2005.

[16] Yoav Freund and Robert E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Journ. of Computer and System Sciences, 55(1), 1997, pp.119-139.

[17] R. Schapire, "A brief introduction to boosting," In Proc. of International Joint Conf. on Artificial Intel., 1999, pp. 1401-1405.

[18] P. Viola and M. Jones, "Fast and Robust Classification using Asymmetric AdaBoost and a Detector Cascade," Advances in Neural Information Processing System 14, MIT Press, 2002.

[19] C. Huang H. Ai, B. Wu and S. Lao. Boosting Nested Cascade Detector for Multi-View Face Detection, proc. ICPR 2004.

[20] J- L. Jiang and K.-F. Loe. S-AdaBoost and Pattern Detection in Complex Environment, proc. CVPR 2003.

[21] A. Torralba. Sharing features: efficient boosting procedures for multiclass object detection, proc. CVPR 2004.

[22] C. Rudin and R. E. Schapire. The Dynamics of AdaBoost: Cyclic Behavior and Convergence and Margins. Journal of Machine Learning Research, 5, 2004.