

Real Time Road Signs Classification

Paolo Medici, Claudio Caraffi, Elena Cardarelli, Pier Paolo Porta
VisLab – Dipartimento di Ingegneria dell'Informazione
Università degli Studi di Parma, ITALY
www.vislab.it
{medici, caraffi, cardar, porta}@vislab.it

Guido Ghisio
Magnetis Marelli Electronics Systems
Strategic Innovation & IP Mgt.
Venaria Reale (Torino), ITALY
guido.ghisio@mmarelli-se.com

Abstract— This paper describes a method for classifying road signs based on a single color camera mounted on a moving vehicle. The main focus will be on the final neural network based classification stage of the candidates provided by an existing traffic sign detection algorithm. Great attention is paid to image preprocessing in order to provide a more simple and clear input to the network: candidate color images are cropped and converted to greyscale, then enhanced using a contrast stretching technique; a multi-layer perceptron neural network is then used to provide a matching score with different road sign models. Finally results are filtered using tracking. Benchmarks are presented, showing that the system is able to classify more than 200 different Italian road sign in real-time, with a recognition rate of 80% to 90%.

I. INTRODUCTION

Road sign detection and recognition can be an important aid to the driver, letting him concentrate on driving; such kind of system can remember signs encountered, even those that go unnoticed or neglected, thus reducing the impact of these events on driving comfort, and also decreasing the possibility of related road accidents.

Road signs are designed to be easily readable, with high contrast and saturated colors, and are installed according to a strict regulation; however, environmental light, weather conditions, paint degradation, dirt, shadows and occlusions make automatic traffic sign recognition a challenging task.

The main goal of this paper is to describe the classification stage of a traffic sign recognition system. Since traffic signs follow strict shape formats, the classification system is driven by the information provided by the shape detection stage, whose output is used to route the input pattern to a specialized classifier.

Several road sign classification techniques are described in literature. One of the simplest methods is cross-correlation with models: in [8] model signs, resampled to 16×16 pixels and rotated, are used to find the best match. Random forests, an ensemble learning technique, are used by [6] to classify signs, and a comparison is made between this technique and SVM and AdaBoost. Support vector machines (SVM) are largely adopted to classify the inner part of road signs. Linear SVM and SVM with Gaussian Kernels are used to recognize the symbol contained in the resampled inner part of road signs: only significant pixels inside the region are used to train the SVM, and each object is only compared with signs with the same shape and color [7]. Since large



Fig. 1. Example of the system output.

inputs are hard to analyze, can be useful for a classifier to reduce its input. Principal component analysis (PCA) and linear discriminant analysis (LDA) techniques can be used to fit this task [5]. Neural networks are also largely adopted, and this technique is also the one chosen to provide the classification stage described and evaluated in this paper. A comparative study between networks with one or two hidden layers has already been made [4], demonstrating that better performance can be achieved using networks with two hidden layers. Tests are also available on the use of Resilient Back-propagation or Scale Conjugate Gradient to train neural networks [9]. Neural networks have also been recently used in embedded systems for traffic sign recognition [2]. Tests have been made on how to train neural networks, using both synthetic and real images. Since a large road signs database can be easily collected, this paper presents exhausting benchmarks to provide a tested and effective indication on how to train neural network for a road sign classification system.

The paper is organized as follows: section II presents the system architecture; section III briefly explains the detection stage presented in a previous article. In section IV is discussed the classification stage. Finally, in sections V and VI, the experimental results and conclusion are reported.

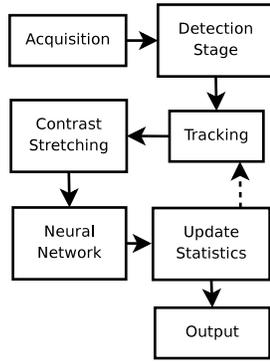


Fig. 2. Traffic Sign Recognition System flowchart.



Fig. 3. Examples of region candidates reported by detection stage.

II. SYSTEM OVERVIEW

The overall system architecture is shown in fig. 2. First, the input image is fed into a detection module in order to extract several regions of interest; a detailed description of the algorithm used can be found in [3], but will be briefly summarized in following section. A tracking stage then associates currently detected regions with previously classified ones; a contrast stretching algorithm tries to compensate for the variable illumination conditions and finally the classifier processes the regions and identifies the sign type. An example of the final output is shown in fig. 1.

III. DETECTION

The detection algorithm is based color and shape analysis, and generates a list of candidate regions where potential road signs can be found.

Color is one of the most exploited features for traffic sign detection, since they usually are characterized by a limited set of easily identifiable colors. This approach thus results in reduced computational requirements. In order to be independent of lighting conditions a color equalization stage is performed on the input image, and the resulting RGB image is converted to the HSV color space, which has been chosen since the Hue component is invariant to changes in light intensity. A threshold on Hue and Saturation is used to obtain Red, Blue, Yellow and White binary images. Binarized images are subsequently labeled, then small labels are rejected, while overlapping ones are merged, and the results are classified using both pattern matching and shape estimation. Example detected candidate regions are shown in fig. 3.

After this stage, each region is assigned to one of eight sign classes: Danger, Temporary Danger, Prohibitions, Obligation, Information, Stop sign, No Parking, Yield and Priority Road.

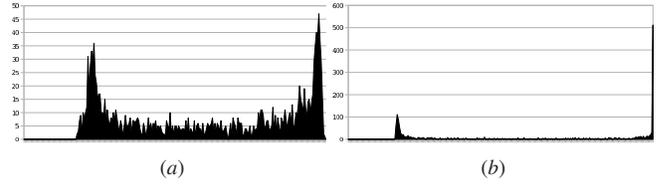


Fig. 4. Illumination histogram of region provided by detection stage, before (a) and after (b) contrast stretching.

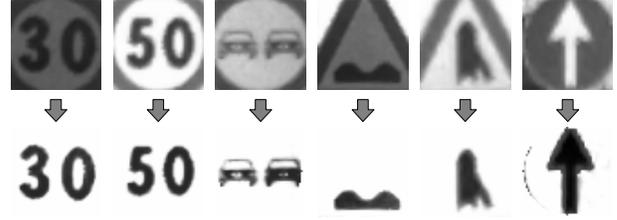


Fig. 5. Contrast Stretching and filtering applied to some test regions provided by detection stage. Illumination variations are rejected and border of sign partially removed.

Signs then undergo a classification stage, with the exception of Yield and Priority Road, since they do not contain particular patterns and are handled separately.

IV. CLASSIFICATION

Regions provided by the sign detector are converted to greyscale and resampled to 50×50 pixels. Since on this regions intensity histograms will be analyzed, and nearest interpolation can generate an imprecise histogram, bilinear interpolation becomes necessary during the resampling process.

In road signs composed by a border and an inner symbol (Danger and Prohibition), only the internal region is resampled to the target size, while the other signs are resampled in their entirety. A mask is then applied on the resampled regions in order to remove the background and the border where present. This mask is generated using shape information (for example a circular mask for Prohibitions and Obligation) and adapted to the border.

From this filtered region a 256 bin greyscale histogram is computed. Since a mask has been applied, the number of pixels analyzed is reduced by up to 50%, but it is still large enough to have a reliable histogram without a filtering pass. On this histogram two peaks are selected, one relative to the background and one relative to the shade of the symbol. For road signs with white inner symbol (Obligation), pixel values are inverted, in order to represent all symbols with a black shade and white background, independently of their class. Different sign classes have different ratios between background and inner symbol size, therefore depending on the specific sign class a different peak detector is used to find the shade indexes for the background and the symbol.

The two detected peaks are used to contrast stretch the filtered region between 40 and 255 (see fig. 4), with the bottom value (40) chosen according to the black level

of synthetic test images. Some examples of refined and illumination corrected regions are shown in fig. 5.

The presented algorithm allows to compensate for scale, position and intensity variations, while noise due to occlusions, dirt or other environmental conditions are not handled by the presented processing steps.

The final reduced and normalized regions of any traffic sign candidate are directly used as the input vector of a neural network classifier, which is presented in the following section.

A. Neural Network

A multi-layer perceptron network with feed forward topology is used for classification [1], and six different neural networks have been trained, one for each reported class of road signs. No special network output is reserved for false positives, since they are expected to provide low values on all output nodes. Networks are trained with the back-propagation method, and validation sets are used to avoid overfitting.

Neural network topologies differ among classes. The input layer is always 50×50 large, and output layer size is equal to the number of signs associated with the given category. The number of hidden layers (one or two) and their size have been chosen to optimize the network performance after extensive benchmarks: the use of only a few neurons and a hidden layer usually does not give satisfactory results, while too many of them causes overfitting. Neural networks with different topologies (with different numbers of layers and layer sizes) have been trained and tested. The best network for each class has been adopted for use in the final application. Currently adopted topologies are shown in table I.

Net Name	# Hidden Layer	Geometry	# Output
Prohibitions	2	115+65	37
Information	1	50	10
Obligation	2	90+40	26
Danger	1	175	42
No-Parking	1	40	3
Stop	1	80	3

TABLE I
NEURAL NETWORK GEOMETRY

A back propagation approach is used to train the networks, using both synthetic patterns (roto-translated synthetic signs) and real ones, manually chosen in order to cover a broad range of cases, but trying to avoid excessive specialization of the network: in order to avoid that kind of specialization, to train the networks only a portion (about 5% 10%) of real signs is used,

The neural networks are functions with 2500 (50×50) inputs and several outputs; each output represents the confidence of classification of the associated sign. Outputs of the network are further processed during the tracking stage described in the following section.

B. Tracking

the use of a tracking stage is important for several reasons:

Set Name	# Pattern
Synthetic	1253
Good Quality	22119
Poor Quality	2467
False Positive	4381

TABLE II
TRAINING/VALIDATION SET

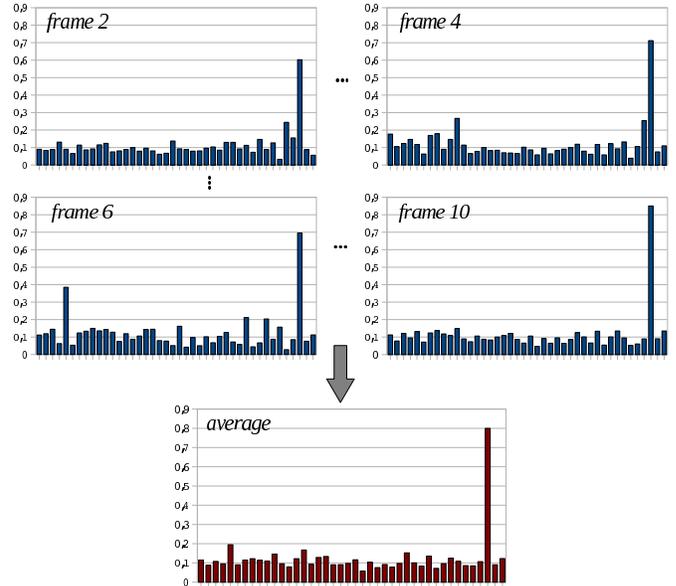


Fig. 6. Neural Network output averaging. Some outputs of neural network of the same sign tracked among 10 frames are showed, and average of all histograms are reported.

- neural networks output is not stable, and can occasionally provide wrong results;
- in order to provide a useful road sign classification system, only one output for each encountered sign should be reported.

Before classification all the regions selected by the detection stage are compared with those classified in the previous frames. For this reason only successfully classified regions can be tracked.

In case of a successful match regions are tracked, while signs remaining without a match are marked as ‘ghost’.

Neural networks can be considered as transfer functions with several inputs, and n outputs $O_0 \dots O_n$. In any frame, outputs of the neural network are averaged with the previous network output $A_0 \dots A_n$ of the same tracked sign, using as weight the region size w , so that far candidates have less weight compared to the closer ones.

$$\begin{aligned} A'_i &= A_i + O_i * w \\ W' &= W + w \end{aligned} \quad (1)$$

When a tracked object goes outside of the screen or becomes ‘ghost’ for more then 5 frames, the sign identifier is reported using the maximum average value

$$id = \max_i A_i \quad (2)$$



Fig. 7. Example of training and validation images: Synthetic signs (a), Good Quality signs (b), Poor Quality signs (c).

only if the normalized output value A_{id}/W is above a given threshold ξ , which in general assumes different values for each signs class. Thresholds have been chosen through the use of receiver operating characteristic (ROC) curves analysis.

In order to reduce the system latency, if a road sign is constantly tracked for at least 10 frames, it is classified in advance and directly reported to the user.

Typical response of a neural network processing the same sign among different frames is shown in fig. 6. The last histogram in the figure is the averaging of all the responses, filtering out most of the outputs variability.

V. EXPERIMENTS AND BENCHMARKS

In this section several results produced during the training stage are reported. Results are presented in the form of ROC curves, where correct detection ratio is a function of the false positives number. ROC curves graphically describe the performance of various classifiers, allowing to compare them without the need of explicitly define a value for the threshold ξ . During network evaluation false positives ratios are computed using both regions wrongly classified to be a sign, and misclassified signs.

All benchmarks presented in this section are produced using the ‘Prohibition Set’, but the following considerations can be directly extended to other networks.

Sets involved in training and validation of networks can be divided in

- **Synthetic:** signs generated by rotation and translation of a model;
- **Good Quality:** real signs, correctly detected, cropped and normalized;
- **Poor Quality:** real signs, partially occluded, incorrectly cropped, and in general not human readable;
- **False Positives:** false positives reported by the detection stage.

Real patterns used to populate the sets are regions dumped by the system before the classification stage. Examples of signs used during training and validation are shown in fig. 7.

The Poor Quality Set is composed by partially occluded signs, saturated signs and any sign hard to classify for a human supervisor. Statistics on this set are made without considering false negatives, but only misclassifications as an error. On this set losing a sign is not important, but it is an error mistaking one sign for an other.

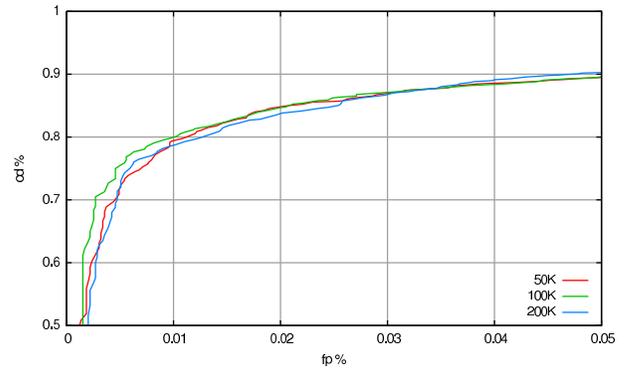


Fig. 8. Overfit on training network: prohibition network with two hidden layers 115+65 validated after 50K (RED), 100K (GREEN), 200K (BLUE) iterations.

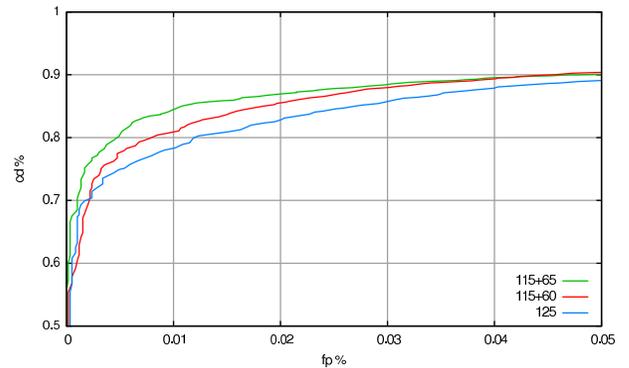


Fig. 9. Prohibition network trained with network of different hidden layers number and size.

Depending on the sign class, the number and quality of false positives changes. False Positives Set also contains real road signs encountered during the system evaluation but not relevant for the user.

A. Neural Network Evaluation

Before any further test, the overfitting problem needs to be discussed, since neural networks trained on a limited training set are likely to overfit. To study this behavior a network has been trained for an high number of iterations. After 50000, 100000, 200000 iterations the network has been evaluated on a validation set, and results are shown in fig. 8. When the network starts to overfit the data, the error on the validation set typically begins to rise.

In order to deal with this problem, the training can be stopped when the validation error steadily increases for a specified number of iterations.

To compare the classification results networks are trained using the same training set, composed by synthetic images and a small portion of real traffic signs. Fig. 9 shows the comparison between one single layer and two different dual layer networks. The network with a first layer of 115 neurons and a second layer of 65 neurons is the best, and it is actually used in the final application.

Initially training has been made using synthetic signs

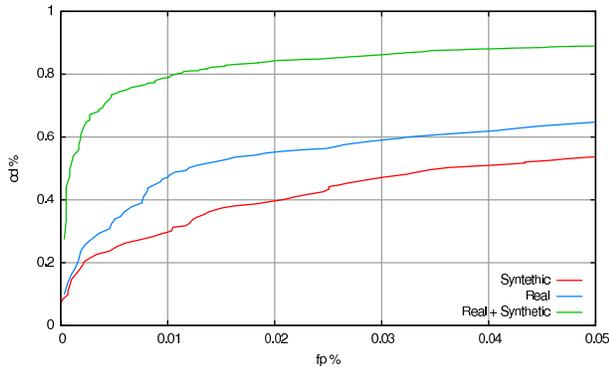


Fig. 10. Prohibition network with two hidden layers 115+65 trained using only synthetic images (RED), only real ones (BLUE) and a mixture of both (GREEN).

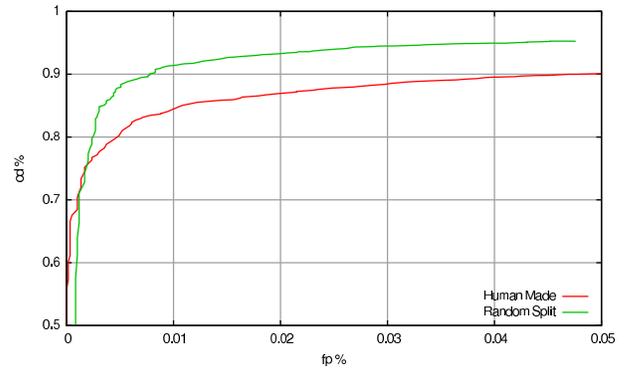


Fig. 12. Prohibition network 115+65 trained using a train set provided by a supervisor (RED) and random generated (GREEN). Both Set approximately contain the same number of patterns.

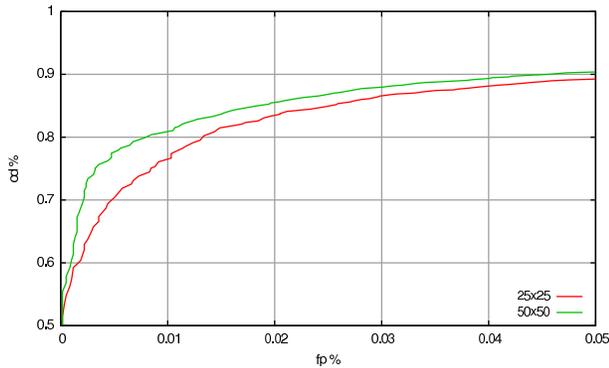


Fig. 11. Prohibition network with two hidden layers 115+60 trained with input size 25×25 pixels (RED) and 50×50 pixels (GREEN).

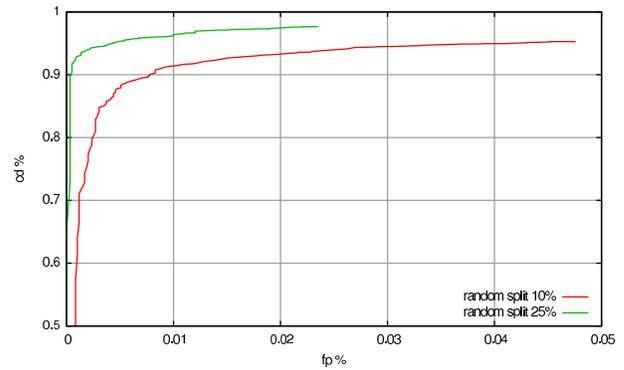


Fig. 13. Prohibition network 115+65 trained using a train random generated using 10% (RED) and 25% (GREEN) of available patterns.

only; however it did not reach the requested performance. Tests with networks trained on synthetic data are reported in fig. 10, with one network trained with synthetic data, one network with real signs and the last with synthetic data and real signs. The training set is composed by 142 synthetic images and 264 real images.

In the works discussed in section I, input patterns sizes range from 16×16 to 31×31 , while patterns collected for this work are larger. The size chosen is 50×50 pixels in order to easily downsample it for evaluation purpose. In fig. 11 comparison between the original network trained on 50×50 pixels samples and one relative to 25×25 downsampled inputs is shown. The 50×50 network is slightly better than the 25×25 , however the last one uses 4 times more memory and in a embedded application this occupation can become relevant.

The construction of training and validation sets used in previous tests has been made by a human, which analyzed the worst classified patterns on the validation set and iteratively moved those patterns from the validation set to the training set. Unfortunately, this procedure has proven to be quite slow, inconsistent and can even decrease the performance of the network, as shown in fig. 12: this test compares a network trained with 623 patterns selected by a human and a network with 696 randomly selected patterns. Approximately both

training set represent 10% of the validation set.

Training set size is still an open issue. Fig. 13 shows the performance of networks trained using 10% and 25% of available signs, however the number of signs in the training set is maintained low on purpose to avoid excessive specialization. In fact, the high amount of collected data appears still limited if compared with the high variability of signs appearance. To improve the performance is thus preferable to increase the number of available signs, instead of increasing only the training set size.

Brief benchmarks of networks with thresholds chosen for the final application are presented in table III.

Net Name	Good Sign CD	Overall CD	Overall FP / Mismatches
Prohibitions	83.3%	79.9%	5.9%
Information	87.9%	97.3%	0.5%
Obligation	94.1%	92.2%	3.2%
Danger	76.4%	76.6%	3.2%
No-Parking	95.1%	94.1%	2.7%
Stop	95.7%	94.5%	1.1%

TABLE III
BRIEF BENCHMARK OF THE CLASSIFICATION STAGE.



Fig. 14. Example of system results in different environment and light conditions. In the bottom part of images the first row shows regions provided by detection stage and the second the output of classification

VI. CONCLUSION

This paper described the final stage of a traffic sign recognition system, capable of classifying in real-time about 200 different Italian road signs, subdivided in six classes, under different light and weather conditions. The system is currently being tested in real environment on a test-vehicle, and it is able to run at rate of 10 Hz, using 752×480 bayer pattern images provided by an off-the-shelf fire-wire camera with a 6 mm lens and a CMOS sensor, mounted behind the windscreen, on a Pentium4 running at 2.0 GHz.

Image regions reported by the detection stage are resampled to 50×50 pixels and illumination normalized. These regions are then fed to a multi-layer neural network, and its output averaged among frames. When a tracked sign disappears the best output is reported as the classification outcome; some results are shown in fig. 14.

ROC curves have been largely used to choose between different approaches and networks topologies in order to globally optimize the system. Classification performance of each network is reported. For normal illumination conditions the system shows optimum performance both in detection and classification. Benchmarks of entire system (detection and classification) made in different daylight conditions on 100 km roads (highway and urban unknown scenarios) shows a correct detection rate of more than 74% of traffic sign.

ACKNOWLEDGMENT

The authors would like to thank Magneti Marelli Electronics Systems for their substantial support to this research.

REFERENCES

- [1] "Lightweight neural network++," <http://lwneuralnetplus.sourceforge.net/>.
- [2] R. Ach, M. Luth, T. Schinner, A. Techmer, and S. Walther, "Classification of traffic signs in real-time on a multi-core processor," in *Proc. IEEE Intelligent Vehicles Symposium 2008*, Eindhoven, Netherlands, June 2008, pp. 313–318.
- [3] A. Broggi, P. Cerri, P. Medici, P. Porta, and G. Ghisio, "Real time road signs recognition," *Intelligent Vehicles Symposium, 2007 IEEE*, pp. 981–986, June 2007.
- [4] A. de la Escalera, L. Moreno, M. Salichs, and J. Armingol, "Road traffic sign detection and classification," *Industrial Electronics, IEEE Transactions on*, vol. 44, no. 6, pp. 848–859, Dec 1997.
- [5] H. Fleyeh, M. Dougherty, D. Aenugula, and S. Baddam, "Invariant road sign recognition with fuzzy artmap and zernike moments," *Intelligent Vehicles Symposium, 2007 IEEE*, pp. 31–36, June 2007.
- [6] A. Z. Kouzani, "Road-sign identification using ensemble learning," in *Intelligent Vehicles Symposium, 2007 IEEE*, 2007, pp. 438–443. [Online]. Available: <http://dx.doi.org/10.1109/IVS.2007.4290154>
- [7] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Siegmann, H. Gomez-Moreno, and F. Acevedo-Rodriguez, "Traffic sign recognition system for inventory purposes," in *Proc. IEEE Intelligent Vehicles Symposium 2008*, Eindhoven, Netherlands, June 2008, pp. 590–595.
- [8] R. Malik, J. Khurshid, and S. Ahmad, "Road sign detection and recognition using colour segmentation, shape analysis and template matching," *Machine Learning and Cybernetics, 2007 International Conference on*, vol. 6, pp. 3556–3560, Aug. 2007.
- [9] N. Yok-Yen and A. Kouzani, "Automatic road sign recognition using neural networks," *Neural Networks, 2006. IJCNN '06. International Joint Conference on*, pp. 3955–3962, 0-0 2006.