

Warning Traffic Sign Recognition using a HOG-based K-d Tree

Fatin Zaklouta and Bogdan Stanciulescu

Abstract—Traffic signs are an essential part of a Driver Assistance System (DAS) and provide drivers with safety information. They are designed to be easily seen and understood. The triangular signs warn the drivers of imminent dangers such as wild animals or a sharp curve. In this paper, an efficient algorithm for the detection and recognition of warning signs is presented. A Histogram of Oriented Gradients (HOG) is used to detect 95% of the triangular warning signs. A blackhat filter eliminates a large part of the false alarms. An approximate nearest neighbors search using a KD-tree refines the result. It eliminates 100% of the remaining false detections and distinguishes amongst the different types of signs. The advantage of using HOG features is that all the warning signs, including static (red frame) and dynamic warning signs (illuminated) can be detected with a single detector and therefore, only one image scan.

I. INTRODUCTION

Traffic signs recognition is an essential part of a Driver Assistance System (DAS) as it provides drivers with safety information. The triangular signs warn the drivers of imminent dangers such as wild animals or a sharp curve. They are designed to be easily seen and understood.

Traffic sign recognition algorithms face three main difficulties: i) the poor quality of the image, due to a low resolution of the camera, bad weather conditions or weak illumination, ii) the rotation, deterioration and size of the traffic signs and iii) the need for processing and memory efficiency in real-time applications such as DAS.

The recognition process consists of a detection and classification step. In the detection step, the image is scanned once and candidate warning signs are retrieved. These are then recognized, in the classification step, as one of the several types of previously learned warning signs or rejected as false alarms. In this paper, an SVM trained with HOG (Histogram of Oriented Gradients) features detects the warning signs. The candidate selection is refined using a blackhat filter and classified by a K-d tree, eliminating false alarms and identifying the traffic signs.

This paper is organized as follows: Section II gives an overview of previous works on traffic sign detection. Section III describes our three-step approach and its components. The results and further analysis are shown in Section IV.

II. RELATED WORK

In the past two decades, many approaches for traffic sign recognition have been proposed. These are mainly based on color and shape features. Color-based methods threshold the

image to obtain the regions of interest. The Hue-Saturation-Value (HSV) color space is one of the most popular since it is more invariant to changes in illumination than RGB. The Hue Saturation Intensity (HSI) is used to detect signs in [14], [12]. It is pointed out in [7], however, that the HSI formula is non-linear and time-consuming to compute. Instead, they propose to use the intensity ratio of the values of one color channel, e.g. red, to the sum of all the RGB components to threshold the image. The resulting regions are classified using a Neural Network. In [2], color-sensitive Haar wavelet features are combined with Adaboost to detect and track traffic signs. More complex approaches include CIECAM97 [9], which is a color appearance standard based on the human perception. The gray-scale based approaches are, however, more robust to changes in illumination than the color-based ones.

Shape-based methods exploit the invariance and symmetry of the traffic signs in the detection phase. In [11], a K-d tree is built with gradient and shape descriptors and used to recognize traffic signs. Franke et al. [8] use Distance Transform (DT) template matching to detect circular and triangular signs. The advantage of matching DTs over edge images is that the similarity measure is smoother. Similarly, [15] use the Color Distance Transform, where a separate distance transform DT is computed for every color channel. The classification is performed using a nearest neighbor template matching approach. In [12], an SVM is trained on the Distance to Border (DtB) vectors to classify the shape of a detected traffic sign. In [10], the FFT of candidate signs are compared to those of the templates. This feature is robust to rotation, deformation and scaling. Speed limit signs are detected using the circular Hough transform [13], [16] and the simpler Radial Symmetry Detector used in [3] are restricted to circular traffic signs.

Many recent approaches use gradient orientation information to detect and classify traffic signs. In [1], Edge Orientation Histograms are computed over class-specific subregions of the image. The Histogram of Oriented Gradients (HOG) [6], initially used for pedestrian detection, have been adapted to traffic sign detection in several works. In [19], [14], the Regions of Interest obtained from color-based segmentation are classified using a HOG-based classifier. To integrate color information in the HOG descriptor, Creusen et al [5] concatenate the HOG descriptors calculated on each of the color channels. To increase the precision of the detector, the SVM detector is iteratively trained and the false alarms from the validation set are reinjected as negatives into the training set.

In our approach, we were interested in training one SVM detector on all types of warning signs, to avoid scanning

Fatin Zaklouta and Bogdan Stanciulescu are with the Robotics Center, Mines ParisTech, 75006 Paris, France
fatin.zaklouta@mines-paristech.fr
bogdan.stanciulescu@mines-paristech.fr

the image several times by each class-specific detector. We also use smaller HOG descriptor vectors, which are less time consuming to compute. Our detector scans an image in 84 to 344 ms depending on its structure. We use gray-scale images to make our detection and classification algorithm more robust to changes in illumination.

III. OUR APPROACH

Our approach is three-step. In the detection phase, an SVM detector is trained with HOG features of all the types of warning signs. Therefore, only one image scan is required to detect all the triangular signs. Next, the candidates found are then filtered by using the blackhat filtering step to eliminate false alarms. In the classification step, a K-d tree distinguishes the different types of signs and eliminates the remaining false positives. Fig. 1 illustrates an overview of our three step approach.

A. Detection Step

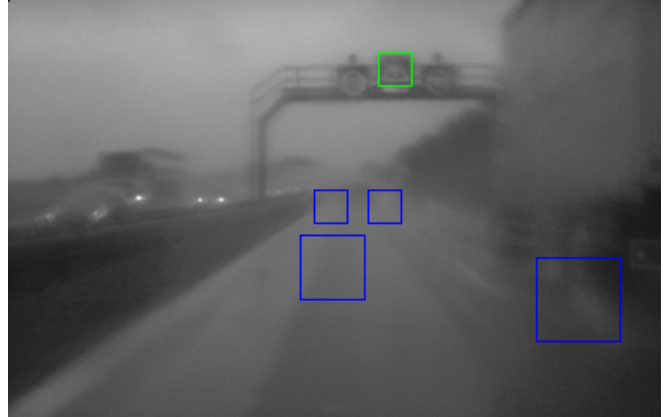
The Histogram of Oriented Gradients, proposed by Dalal and Triggs [6] for pedestrian detection, has become popular for the recognition of different types of objects. We chose to use this feature for the alert sign detection due to its scale-invariance, local contrast normalization, coarse spatial binning and weighted gradient orientations. The unsigned gradients allow for the detection of both static signs as well as dynamic, illuminated signs (see Fig. 2) with a single detector. This is not possible with other features, such as Haar wavelet features [18], where the difference between the sums of pixels in two regions indicates the direction of the gradient.

Fig. 3 illustrates the structure of the HOG descriptor. The image of a traffic sign is divided into overlapping blocks. Each block, in turn, is divided into non-overlapping cells. The gradient magnitude and angle are computed for each pixel. A histogram of the orientations of the gradients of each cell is formed. The magnitude of the gradient is used as a vote weight. The histograms of the cells of each block are concatenated to form the descriptor.

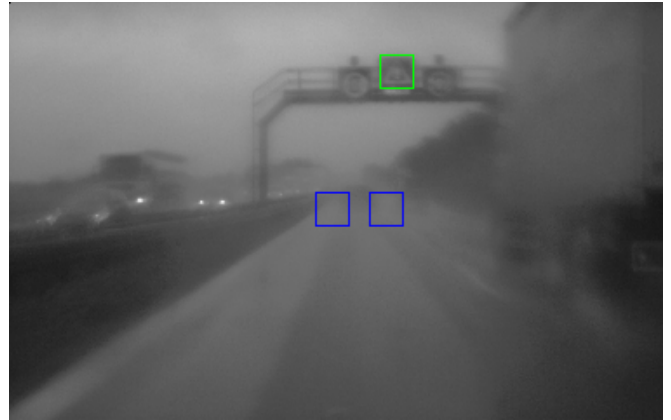
In this paper, the sizes of the cells and blocks are varied to obtain different descriptor sizes, from 144 to 1764 values. We use 9-bin histograms for all our experiments. The sign/non-sign training images are rescaled to 16x16 pixel images. Their HOG descriptors are computed and used to train a linear SVM classifier using the SVMLight¹ library. The n resulting support vectors are combined to a single one by multiplying each vector by its weight α and adding it to a global vector. The weight α_i is the confidence in support vector v_i and is computed as a function of its classification error.

$$v = \sum_i^n \alpha_i \cdot v_i$$

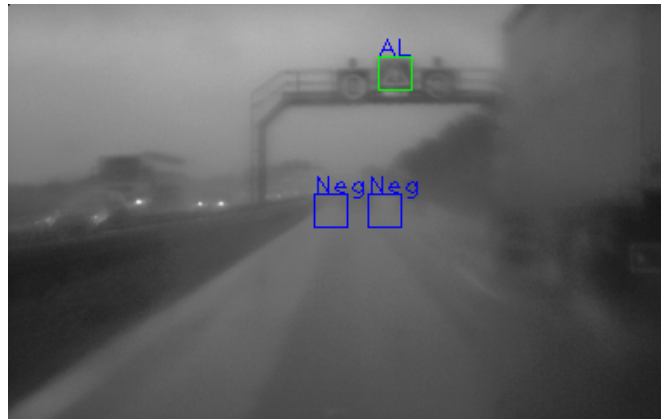
This reduction to one support vector accelerates the detection as the HOG descriptors of the candidates are compared to a single vector instead of several. During the detection phase,



(a) Initial detection using SVM detector with 576 value HOG descriptor



(b) Blackhat filtering removes some false alarms



(c) Classification using K-d tree recognizes the illuminated alert warning sign (AL) and identifies the remaining false alarms (Neg)

Fig. 1. Three step approach: a) Detection using SVM detector, b) Blackhat filtering and c) K-d tree classification

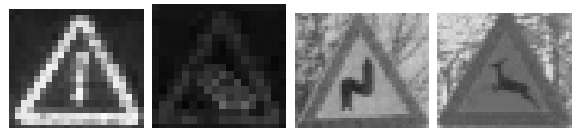


Fig. 2. Examples for dynamic (illuminated) and static alert signs

¹<http://svmlight.joachims.org/>

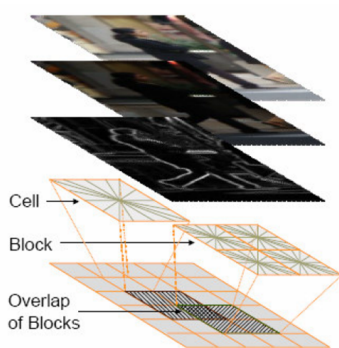


Fig. 3. Structure of HOG descriptor [6]

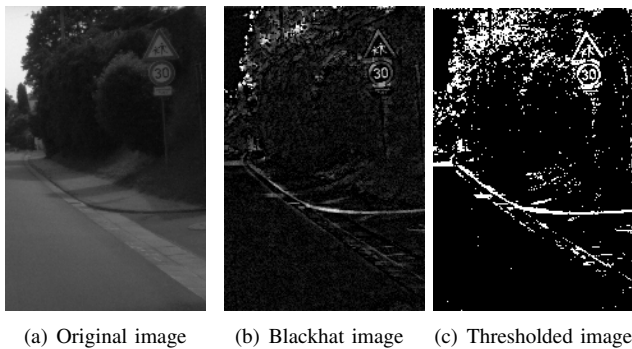


Fig. 4. Image segmentation using blackhat

the image is scanned at multiple scales, starting with a 16x16 pixel window. The resulting candidates are passed on to the K-d tree for further refinement and classification.

B. Blackhat Filtering

To improve the precision of the SVM detector, we filter the detected candidates using a morphological operator. A blackhat (bottom-hat) transform of the image with a window size of 3x3 pixels is applied to the image. The blackhat transform is defined as the difference between the closing and the input image. This morphological filter emphasizes dark pixels with a high contrast to their local environment, which is the case for traffic signs. The resulting image is then thresholded. This threshold corresponds to twice the mean over the whole image. The mean over the entire image helps take the global illumination into consideration, while a mean computed locally would be more sensitive to small illumination variations in the image. The result is a segmentation of the dark regions that are surrounded by light pixels. Fig. 4 shows the results of the blackhat segmentation. Note that the blackhat filter emphasizes the border and the pictogram of the traffic sign as well as the border of the sidewalk and the light passing through the trees. However, a large part of the image is eliminated using this filter and the number of false alarms is reduced.

C. Classification Step

The K-d tree is a nearest-neighbor-based search tree. We use the Best Bin First algorithm [4] to perform an approximate Nearest Neighbor search. This efficient variant

of the K-d tree search algorithm allows for the indexing of higher dimensional spaces which is required when using the large HOG vectors. Instead of exhaustively searching the entire tree, Beis et Lowe [4] propose to search the nodes closest to the query using a priority queue. This queue contains candidate nodes ranked according to their distance to the query. This ranking determines the order in which the nodes are examined. The search process iteratively adds the siblings of the current node being examined into the priority queue. The search is terminated when the algorithm scans a predefined maximum number of nodes E_{\max} . An $E_{\max} = 400$ achieved the best results in our experiments.

A nearest neighbor approach is more suitable in our case due to the small amount of training samples for some alert sign classes. Training a learning-based classifier such as an SVM to a highly imbalanced dataset often requires parameter fine-tuning, making the system less generic.

During the training, the images in the training set are scanned by the SVM detector. The resulting true and false positive windows extracted are used to build the K-d tree. When testing, the detected candidates are searched for in the K-d tree to identify false alarms and recognize the found traffic signs. For each candidate, the k nearest neighbors (kNN) are retrieved. The vote of the class, to which each of the nearest neighbours belongs is incremented. The reciprocal of the Euclidean distance to the candidate is used as a vote weight. The maximum vote determines the class of the candidate.

$$\text{vote}_{\text{class}} = \sum_1^k \frac{1}{\text{distance}(kNN, \text{candidate})}, kNN \in c$$

IV. RESULTS AND ANALYSES

We evaluate the proposed detection and classification techniques using real-world image sequences taken in both urban and highway environments under different weather conditions. Therefore, the traffic sign images include varying illumination, scale and pose. The images are down-scaled to 376x240 pixels to accelerate the scan.

The recall and precision measures used to evaluate the performance of our algorithm are defined as follows:

$$\text{recall} = \frac{\text{true positives detected (TP)}}{\text{total true positives}}$$

$$\text{precision} = \frac{\text{true positives detected (TP)}}{\text{all detections}}$$

A. Data set

Our data sets contains 11 kinds of traffic signs. Their sizes range from 15 to 52 pixels. The non-sign examples of 25x25 pixels are extracted at random from the training images. The images are taken under different weather and illumination conditions and partially deteriorated signs. Examples from each class are illustrated in Fig. 5. Table I shows the distribution of the 11 types of classes and the train/test ratio. Note that the dataset is unbalanced. The dynamic warning signs make up to 42 % of all the signs while the proportion of the dynamic traffic jam lies at only 1 %. This imbalance

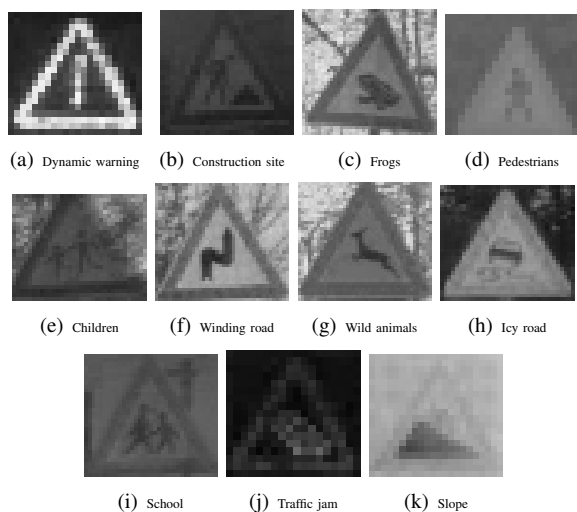


Fig. 5. Examples of the 11 traffic sign classes in our data set

TABLE I
NUMBER OF TRAINING AND TESTING SIGNS

Sign	Total	Train	Test	Proportion
Dynamic Warning	342	309	33	42 %
Construction site	15	13	2	2 %
Frogs	17	15	2	2 %
Children	59	53	6	7 %
Winding road	79	71	8	10 %
Pedestrians	37	33	4	5 %
Wild animals	17	17	2	2 %
Icy road	190	171	19	23 %
School	19	17	2	2 %
Dynamic traffic jam	12	10	2	1 %
Slope	33	30	3	4 %

of the classes favors a generic nearest neighbors approach using the K-d tree as opposed to a learning algorithm.

B. Detection

A single binary SVM detector was trained using 737 traffic signs of all types (see Table I) and an equal number of non-sign images, chosen at random from the train images. The block and cell sizes and the stride of the HOG descriptor were varied. The HOG descriptors of 144 to 1764 values were trained and tested. We used the criteria applied in the Pascal VOC challenge to determine true positives and to compare the ground truth and the estimated bounding box based on the ratio of their intersection and the area of their union.

$$\frac{candidate \cap ground\ truth}{candidate \cup ground\ truth} > 50\% \Rightarrow candidate \in TP$$

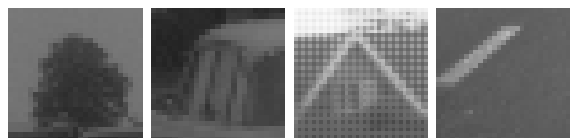


Fig. 6. False alarms include tree tops, roofs, car sides and road markings



Fig. 7. Examples of detections using the 576 value HOG detector. Our dataset contains images taken in urban, country road and highway environments under different illumination and weather conditions.

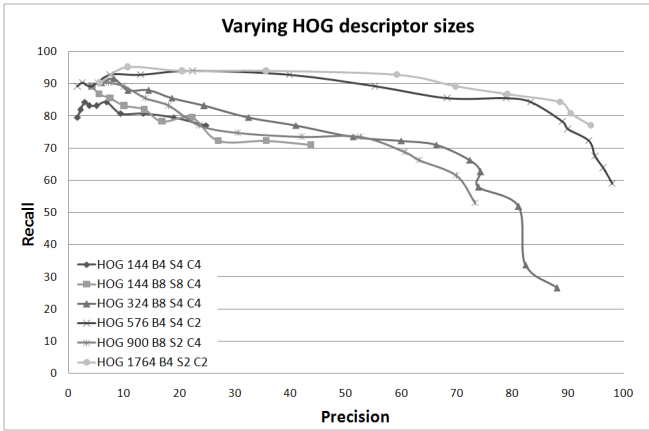


Fig. 8. Recall and precision of HOG detectors, sizes 324, 576, 900, 1764. B = Block size, S = Slide size, C = Cell size in pixels

Fig. 8 illustrates the recall/precision curves of the different size HOG descriptors. The smaller cell and block sizes achieve higher recall and precision rates due to the finer spatial binning of the gradient orientations. In particular, the HOG descriptors with 576 and 1764 values performed the best with a highest recall rate of 95 %. Examples of the detection results of the 1764 value HOG detector are shown in Fig. 7. The former achieved a recall of 87 % with a precision of 79 %, the latter a recall of 89 % and precision of 70 %. Many of the false alarms are due to tree tops, roofs, car sides and road markings as shown in Fig. 6.

The speed of the SVM detector depends on the block, cell and stride sizes of the HOG descriptor used. The 144 value detector takes 84 ms per image on average, the 324 value detector about 103 ms and the 900 value detector takes 263 ms. The detectors with smaller cell and block sizes, the 576 and 1764 value detectors take 344 and 247 ms respectively to detect the traffic signs in an image since the finer spatial binning has a longer computation time. The tests were run on a Core 2 laptop, with 2.39 GHz and 3.5 GB RAM.

C. Blackhat Filtering

As described in Section III-B, the number of false alarms is reduced using a black hat transform. The result of this false alarm filter is illustrated in Fig. 9. The precision is almost doubled, while the recall deteriorates only slightly. For example, a 900 value detector with a recall of 94 % and a precision of 22 % achieves 91 % and 33% respectively after the blackhat filtering.

D. Classification

The candidates extracted by the SVM detector are passed on to the K-d tree classifier to identify the traffic sign type and eliminate false alarms. For each candidate, the k nearest neighbors in the K-d tree are found and the class vote is inversely proportional to the distance between the candidate and its closest match (see Section III-C).

Table II shows the classification results obtained using different size HOG descriptors. About 94-100% of the false alarms are eliminated using any of the HOG descriptors. This

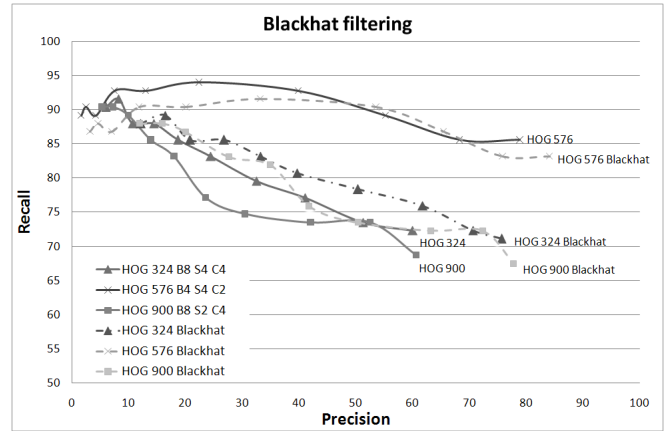


Fig. 9. Recall and precision of HOG detectors with and without blackhat filtering, sizes 324, 576, 900. B = Block size, S = Slide size, C = Cell size in pixels

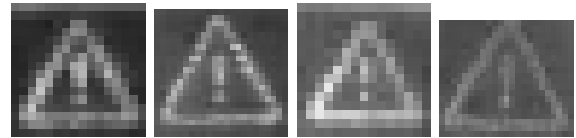


Fig. 10. Changing illumination intensity makes it more difficult to classify the dynamic traffic signs

is partially due to the abundance of the false alarms in the training set of the K-d tree. Most traffic sign classes, such as the *Pedestrians* and the *Icy Road* yield high classification rates of 100% and 85.7%. A large part of the dynamic warning signs are misclassified, due to the changing intensity of the signs' illumination as shown in Fig. 10.

1) *Competition: German Traffic Sign Recognition Benchmark*: Since our dataset is relatively small, we further test the performance of the HOG features with the K-d trees using the German Traffic Sign Recognition Benchmark [17]. This image data set contains 43 classes, 26,640 train images and 12,569 test images. The three HOG descriptors provided by the challenge with 1568, 1568 and 2916 values are used.

The performance of the K-d tree improves when increasing

TABLE II

CLASSIFICATION RATES OF 11 TYPES OF TRAFFIC SIGNS AND BACKGROUND WITH VARYING HOG DESCRIPTOR SIZES

HOG Descriptor	Number	324	576	900
Dynamic Warning	33	54.3 %	78.1 %	55.9 %
Construction site	2	100 %	0 %	100 %
Frogs	2	100 %	100 %	100 %
Children	6	60 %	50 %	50 %
Winding	8	88.9 %	50 %	100 %
Pedestrians	4	100 %	100 %	100 %
Wild	2	100 %	50 %	100 %
Icy	19	57 %	75 %	85.7 %
School	2	50 %	0 %	50 %
Traffic jam	2	0 %	100 %	0 %
Slope	3	75 %	100 %	100 %
Background		99.6 %	94 %	99.9 %
Average		73.7 %	66.4 %	78.5 %

TABLE III

CLASSIFICATION RESULTS WHEN VARYING E_{\max} IN K-D TREE

E_{\max}	HOG 1	HOG 2	HOG 3	HOG 4	Avg ms
500	71,30 %	68,84 %	71,08 %	90,69 %	1.75
1000	73,03 %	70,40 %	72,75 %	91,44 %	3.58
1500	73,76 %	71,25 %	73,65 %	92,00 %	5.37
2000	73,94 %	71,87 %	74,00 %	92,29 %	7.16
2500	74,37 %	72,03 %	74,53 %	92,36 %	8.95
5000	74,92 %	73,39 %	75,03 %	92,70 %	17.9

the E_{\max} variable yet the processing time is longer. The best classification rate achieved is 92.7% at 17.9 ms per image on average. The detailed results are given in [20]. The Random Forests achieve 97.2% yet require longer computation times.

V. CONCLUSIONS AND FUTURE WORKS

In this paper, we propose a warning traffic sign detection approach based on HOG feature learning and an approximate nearest neighbor classification using K-d trees. A single SVM detector, trained on all warning signs, is used to scan the image only once to detect all types of signs. The experiments performed show that a this detector has a recall of 95 % for the 11 types of traffic signs. It takes 84 ms (144 values) to 344 ms (576 values) to scan an image depending on the granularity of the spatial binning of the gradient orientations in the HOG descriptor, i.e. its size. The blackhat filtering step doubles the precision by segmenting the regions of interest in the image. The K-d tree classifier eliminates the remaining false alarms and classifies the candidates found at an average rate of 78.5 %. Our algorithm was evaluated on a dataset containing real world images taken under varying weather and illumination conditions in both urban and highway environments.

In future work, the detection will be extended to include other shapes of traffic signs. The classification rates can be improved by enriching the data set and by exploiting other visual information such as the color and distance transform features. The system performance can also be improved by integrating spatial and temporal information i.e. restricting the search space to the sides of the road and reinforcing the classification decision over several consecutive frames.

REFERENCES

- [1] B. Alefs, G. Eschemann, H. Ramoser, and C. Beleznaï. Road sign detection from edge orientation histograms. In *2007 IEEE Intelligent Vehicles Symposium*, pages 993–998, June 2007.
- [2] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler. A system for traffic sign detection, tracking, and recognition using color, shape, and motion information. In *Proceedings of the IEEE Symposium on Intelligent Vehicles*, pages 255–260. Citeseer, 2005.
- [3] N. Barnes, A. Zelinsky, and L.S. Fletcher. Real-time speed sign detection using the radial symmetry detector. *Intelligent Transportation Systems, IEEE Transactions on*, 9(2):322–332, 2008.
- [4] J.S. Beis and D.G. Lowe. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 1000–1006. IEEE, 2002.
- [5] I.M. Creusen, R.G.J. Wijnhoven, E. Herbschleb, and P.H.N. de With. Color exploitation in hog-based traffic sign detection. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 2669–2672, 2010.
- [6] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, editors, *International Conference on Computer Vision & Pattern Recognition*, volume 2, pages 886–893, INRIA Rhône-Alpes, ZIRST-655, av. de l’Europe, Montbonnot-38334, June 2005.
- [7] A. De La Escalera, L.E. Moreno, M.A. Salichs, and J.M. Armingol. Road traffic sign detection and classification. *IEEE Transactions on Industrial Electronics*, 44(6):848–859, 1997.
- [8] U. Franke, D. Gavrilă, S. Görzig, F. Lindner, F. Paetzold, and C. Wöhler. Autonomous driving approaches downtown. *IEEE Intelligent Systems*, 13(6):1–14, 1999.
- [9] X.W. 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*, 17(4):675–685, 2006.
- [10] P. Gil-Jimenez, S. Lafuente-Arroyo, H. Gomez-Moreno, F. Lopez-Ferreras, and S. Maldonado-Bascon. Traffic sign shape classification evaluation. Part II. FFT applied to the signature of blobs. In *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*, pages 607–612. IEEE, 2005.
- [11] Wen-Jia Kuo and Chien-Chung Lin. Two-stage road sign detection and recognition. In *Multimedia and Expo, 2007 IEEE International Conference on*, pages 1427–1430, 2007.
- [12] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, and F. Lopez-Ferreras. Road-sign detection and recognition based on support vector machines. *IEEE transactions on intelligent transportation systems*, 8(2):264–278, 2007.
- [13] F. Moutarde, A. Bargeton, A. Herbin, and L. Chanussot. Robust on-vehicle real-time visual detection of American and European speed limit signs, with a modular Traffic Signs Recognition system. In *2007 IEEE Intelligent Vehicles Symposium*, pages 1122–1126, 2007.
- [14] X. Qingsong, S. Juan, and L. Tiantian. A detection and recognition method for prohibition traffic signs. In *Image Analysis and Signal Processing (IASP), 2010 International Conference on*, pages 583–586. IEEE, 2010.
- [15] A. Ruta, Y. Li, and X. Liu. Real-time traffic sign recognition from video by class-specific discriminative features. *Pattern Recognition*, 43(1):416–430, 2010.
- [16] A. Ruta, Y. Li, M. Uxbridge, F. Porikli, S. Watanabe, H. Kage, K. Sumi, and J. Amagasaki. A New Approach for In-Vehicle Camera Traffic Sign Detection and Recognition. In *Proc. IAPR Conference on Machine Vision Applications, Japan*, 2009.
- [17] Johannes Stallkamp, Marc Schlipsing, Jan Salmen, and Christian Igel. The German Traffic Sign Recognition Benchmark: A multi-class classification competition. In *International Joint Conference on Neural Networks*, 2011.
- [18] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE CVPR*, volume 1, pages 511–518, 2001.
- [19] Y. Xie, L.F. Liu, C.H. Li, and Y.Y. Qu. Unifying visual saliency with HOG feature learning for traffic sign detection. In *Intelligent Vehicles Symposium, 2009 IEEE*, pages 24–29. IEEE, 2009.
- [20] Fatin Zaklouta, Bogdan Stanculescu, and Omar Hadmoun. Traffic sign classification using k-d trees and random forests. In *IJCNN*, 2011.