Large-Scale Traffic Sign Recognition based on Local Features and Color Segmentation

M. Blauth, E. Kraft, F. Hirschenberger, M. Böhm

Fraunhofer Institute for Industrial Mathematics,
Fraunhofer-Platz 1, D-67663 Kaiserslautern
University of Applied Sciences Kaiserslautern, D-67657
Morlauterer Strasse 31
eMail: marco_blauth@web.de
eMail: erwin.kraft@itwm.fraunhofer.de
eMail: f.hirschenberger@fh-kl.de
eMail: martin.boehm@fh-kl.de

Abstract

We present a state-of-the-art traffic sign recognition system that is able to detect and classify 43 different German traffic signs. To maintain a high performance and scalability of the system we use a two-stage approach that separates the detection from the classification step. We use a fast shape detection method based on multi block local binary patterns (MB-LBP) [1], which has shown to give good results for a face detection application. Color segmentation is also used alongside the shape detector. This increases the affine invariance properties and overall detection rate of the system. For the classification, we rely on HOG-Features [2] and nonlinear support vector machines (SVMs). Our classifier achieves a score of 96.93% on the German Traffic Sign Recognition Benchmark [3], which can be considered as a top scoring result. Since this benchmark is a pure classification benchmark that uses bounding boxes with centered signs we also evaluated the detection capabilities of our system on 108 traffic images provided by a street inspection company. Our system detected and classified 97% of the traffic signs in this data set correctly while the false positive rate was very low. The results show a high applicability of our system under real world conditions.

1 Introduction

Reliable traffic sign recognition can be considered as a key aspect of driver assistance systems. Most commercially available systems are limited to a subset of important traffic signs like speed limitation or give way signs that are detected from a frontal view. In this paper we deal with the problem of large scale traffic sign recognition where we have to process a wide variety of different traffic sign classes. Our motivation to the problem are driven by the needs of a street inspection company: Cameras are mounted on inspection cars and images of the road scene are taken every four meters while the car moves along. Our goal is to find all traffic signs in the scene. The cameras observe the scene from different viewing angles, which means that the traffic signs are more likely to be distorted as it would be the case with frontal views. Our system has to differentiate between visually very similar traffic signs, for example different speed limitation signs. It also has to be robust under real world conditions, such as occlusions, image blur, varying lighting conditions as well as other distortions. Some examples are shown in figure 1.

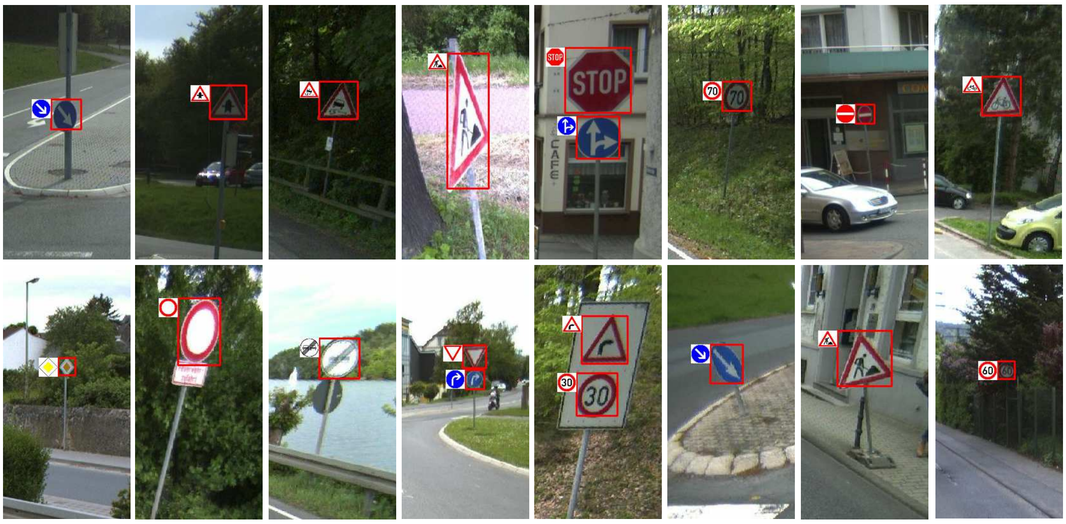

Figure 1: Successful detection and classification of different traffic signs. Our system is able to discriminate between visually similar signs and has good scale and affine invariance properties.

The literature shows numerous works on the detection and classification of traffic signs. Since there are many published methods that handle the processing of a limited number of important signs well the area of interest has shifted to large-scale settings with more difficult environmental conditions. The recently published German Traffic Sign Recognition Benchmark (GTSRB) [3] was designed to address the classification problem in such a case. Earlier published methods like [4], [5], [6] and others do not perform well on this data set since they require a precise segmentation of the signs and text/pictograms, which is often not possible for the low contrast and distorted images in this benchmark. Newer approaches like [7] or [8] use HOG-Features [2], which have shown to give excellent results in difficult detection and classification scenarios. In our experiments they significantly outperformed other features, for example distance-to-border features [4].
HOG-Features [2] are often used by sliding-window detectors that run across various scales where each window is classified by a support vector machine (SVM). From our experience,

such detectors can become a performance bottleneck when dealing with many different classes and nonlinear kernel functions. We therefore use a two-stage approach where we first detect candidate regions in a more efficient way based on shape and color information. These candidates are then classified using HOG-Features [2] to determine whether they contain a traffic sign. Our approach is very fast: It processes a high resolution image of 1388×1038 pixels in under 2 seconds on a normal workstation. Also, it increases the detection and classification rate for distorted traffic signs that are viewed from steep angles.

2 Detection of traffic signs

The goal of our detection stage is to identify image regions that may contain a traffic sign. To ensure a high system performance we focus on fast detection methods. Shape and color information are the primary information cues to detect regions of interest. The found image regions are then normalized and classified using HOG features [2]. Our HOG-classifier is robust against distortions and is also very suitable to reject false positives. Therefore, we focus on a high detection rate both for the shape and color detection rather than a low number of false positives.
We use 108 street images to evaluate our detection results, because the GTSRB [3] is a pure classification benchmark and is not well suited for this purpose. Table 1 shows the detection results on our data set.

Table 1: Detection results for a set of 108 street images.

	Detection rate	Number of false positives
MB-LBPs	97.7%	218
Haar-like features	92%	377
Color segmentation	99%	1396
Color segmentation + MB-LBPs	100%	1396 + 218

2.1 Color based detection

Color segmentation and blob detection are used to find interesting image regions. We look for the color ranges red, blue and yellow in the HSI color space, because they are important traffic sign colors. The HSI color space separates the color information from the greyscale intensity. The color is encoded in the hue (H) and saturation (S) channels while the greyscale intensity (I) uses its own channel. However, color information is not reliable if the saturation is very low. We therefore define intervals $[H_{min}, H_{max}]$, $[S_{min}, S_{max}]$ and $[I_{min}, I_{max}]$ for all three channels and use a simple binary thresholding function:

$$B(i,j) = \begin{cases} 1, & \text{if } (H_{min} \leq H(i,j) < H_{max}) \wedge \\ & \quad (S_{min} \leq S(i,j) < S_{max}) \wedge \\ & \quad (I_{min} \leq I(i,j) < I_{max}) \ , \\ 0, & \text{otherwise} \end{cases}$$

The intervals are chosen by grid search so that the detection rate is maximized for a training set of images. Our color detection results are shown in table 1. As one can see the detection rate is very high but there are also many false positives. Some of the false positives are shown in figure 2.

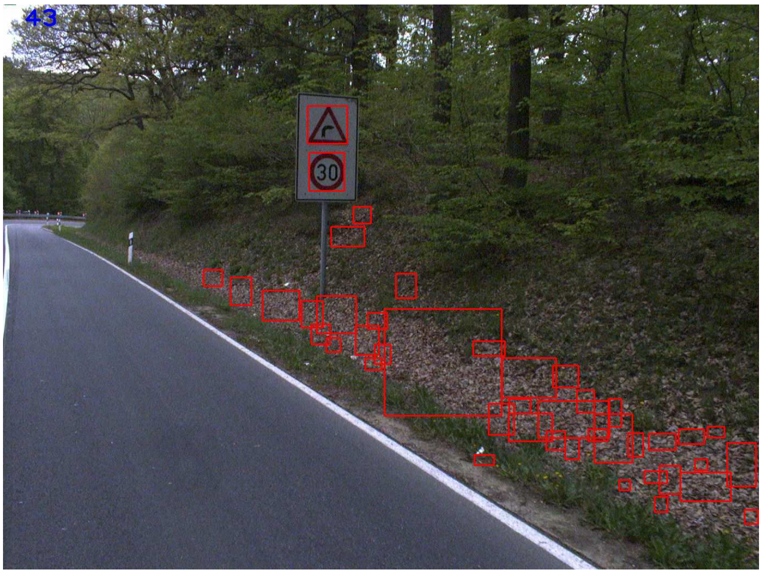

Figure 2: False positive detections of our color segmentation method: Both traffic signs are detected but the red colored leaves on the ground produce a large number of false positives. The false positives are later rejected by our HOG-classifier.

The detected regions are normalized to increase the affine scale invariance properties of our system. An example is shown in figure 3: The detected traffic sign is distorted because of the steep viewing angle. A coarse spatial normalization procedure is sufficient to obtain a correct classification result.

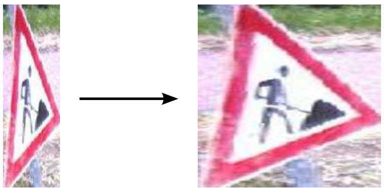

Figure 3: Affine normalization of segmented image regions. The left image is too distorted for our HOG-classifier. After normalization the region could be correctly classified as a valid traffic sign.

2.2 Shape based detection

For the shape based detection of traffic signs we rely on the methods presented in [1] and [9]. The detector uses multi block local binary patterns (MB-LBPs). A MB-LBP is a local descriptor that encodes greyscale differences between local image sub-regions. Figure 4 gives a schematic overview of the computation of MB-LBPs. The computation is very simple and can be effectively speeded up by integral images.
Training is done using a variant of the popular AdaBoost algorithm. AdaBoost combines so called weak classifiers $h_j(x)$ into a strong classifier $h(x)$. A weak classifer is restricted to use only one feature f_j from a predefined set:

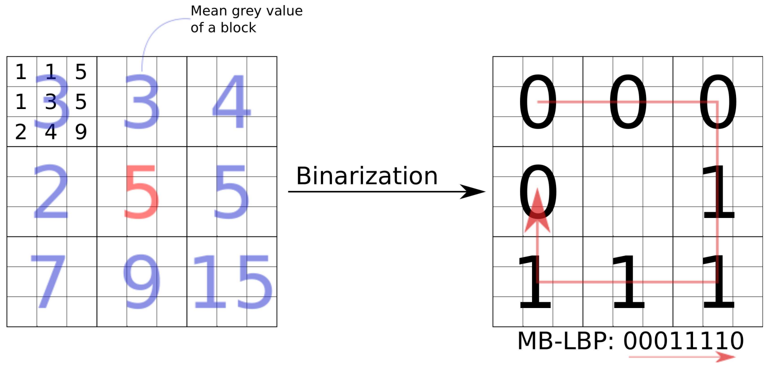

Figure 4: Computation of the MB-LBP descriptor: The mean grey values of the surrounding blocks are compared to the mean value of the center block. The resulting binary pattern forms a compact descriptor.

$$h_j(x) = \begin{cases} 1, & \text{if } p_j f_j(x) \geq p_j \theta_j \\ 0, & \text{otherwise} \end{cases}$$

θ_j is a threshold and p_j a parity. The function $f_j(x)$ applies a feature f_j to an image region x. The final strong classifier $h(x)$ is a combination of weak classifiers $h_t(x)$ that were selected at the training stage t:

$$h(x) = \begin{cases} 1, & \text{if } \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0, & \text{otherwise} \end{cases}$$

where $\alpha_t = \log \frac{1}{\beta_t}$ and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$. ϵ_t is the overall error of the weak classifier $h_t(x)$. Weak classifiers are trained and selected so that they complement each other. The final detector then arranges multiple strong classifiers in a cascade: At each stage a classifier is trained by AdaBoost and the number of allowed features is increased so that classifiers at later stages become more reliable but also more complex than classifiers at early stages of the cascade. The main idea is that image regions that do not contain objects of interest are discarded at early stages of the cascade. Thus, the detector becomes very efficient when used in a sliding-window manner. More details on the training and construction of the cascade can be found in [1] and [9].

We used images from the training set of the German Traffic Sign Recognition Benchmark (GSTBR) to train our shape detector. The detector was able to predict the correct traffic sign shape on the test set with a success rate of 96%. We also tested our detector on our 108 street images. The detection rate on this data set was 97.7% and the number of false positives was 218 (see table 1). We also compared the MB-LBPs with Haar-like features. The MB-LBPs produced a smaller number of false positives while the detection rate was higher. Some detection results and false positives can be seen in figure 5.

3 Classification of traffic signs

In the classification stage we determine if a detected image region contains a particular traffic sign or if it has to be rejected as a false positive. The detected regions are first normalized to the size of 40 × 40 pixels and then HOG features [2] are computed on a

Figure 5: False positive detections of our shape detector: The traffic sign is detected but so are other round objects like tires. The false positives are later rejected by our HOG-classifier.

dense grid. A HOG feature is a local image feature that captures shape and structure information. It is a weighted 3D histogram that quantizes the spatial positions and directions of the image gradients. To avoid binning effects it is very important to use linear interpolation between all the adjacent bins in the 3D histogram. Figure 6 gives an overview of the computation of HOG features. Further details can be found in [2].

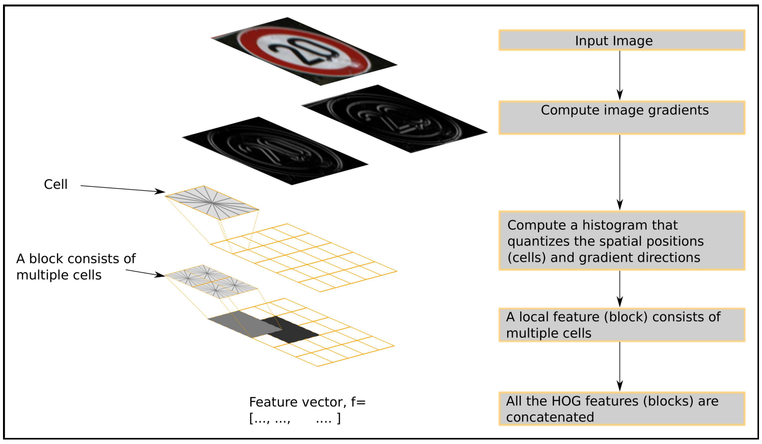

Figure 6: Computation of HOG features: The image region is divided into cells. For each cell the directions of the image gradients are quantized into bins. Linear interpolation is used between adjacent cells and direction bins.

Our HOG features consist of 2×2 cells and each cell covers an area of 5×5 pixels. All HOG features are concatenated to form a high dimensional feature vector that describes the whole image region. These feature vectors are used as inputs for non-linear support vector machines (SVMs), which are trained in a one-vs-one manner so that they are able to classify 43 different traffic signs. The HOG-classifier was trained using the training set of the GTSRB [3]. On the test set it was able to classify 96.93% of the images correctly, which is a very high success rate. We also tested the classifier on the detected regions of our 108 street images. 97% of the traffic signs were classified correctly while only 44 false positives remained. Some examples are shown in figure 1: The classifier is able to discriminate between very similar signs and reduces the number of false positives from the detection stage significantly.

Table 2: Detection and classification results for a set of 108 street images.

Recognition rate (detection and classification)	Number of false positives
97%	44

We also tried to compute HOG-features on different color channels but found that the classification rate could not be increased that way.

4 Conclusions

A traffic sign detection and recognition system was presented that is able to reliably recognize a large number of traffic signs under real world conditions. The detection and classification capabilities of the system were extensively tested using two different data sets: The German Traffic Sign Recognition Benchmark [3] and a detection data set, which consists of 108 street images. Our HOG-classifier achieved a score of 96.93% on the GTSRB. The overall recognition rate on our detection data set was 97% while there were only 44 false positives.

5 Acknowledgements

This work was supported by the Bundesministerium für Bildung und Forschung (BMBF).

References

[1] S. C. Liao, X. X. Zhu, Z. Lei, L. Zhang, and S. Z. Li. Learning multi-scale block local binary patterns for face recognition. In *International Conference on Biometrics (ICB)*, volume 4642, pages 828–837, 2007.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *International Conference on Computer Vision & Pattern Recognition (CVPR)*, volume 2, pages 886–893, 2005.

[3] German traffic sign recognition benchmark. http://benchmark.ini.rub.de.

[4] S. Lafuente-Arroyo, P. Gil-Jiminéz, R. Maldonado-Bascon, F. López-Ferreras, and S. Maldonado-Bascón. Traffic sign shape classification evaluation i: Svm using distance to borders. In *Proceedings of the 2005 IEEE Intelligent Vehicles Symposium*, 2005.

[5] L-W. Tsai, J.-W. Hsieh, C.H. Chuang, Y.J. Tseng, K.C. Fan, and C.C. Lee. Road sign detection using eigen colour. In *IET Computer Vision*, volume 2, pages 164–177, 2007.

[6] X. Chen, J. Yang, J. Zhang, and A. Waibel. Automatic detection and recognition of signs from natural scenes. In *IEEE Transactions on Image Processing*, volume 13, pages 87–99, 2004.

[7] F. Zaklouta, B. Stanciulescu, and O. Hamdoun. Traffic sign classification using k-d trees and random forests. In *IEEE International Joint Conference on Neural Networks (IJCNN)*, pages 2151–2155, 2011.

[8] I. Bonaci, I. Kusalic, I. Kovacek, Z. Kalafatic, and S. Segvic. Addressing false alarms and localization inaccuracy in traffic sign detection and recognition. In *16th Computer Vision Winter Workshop*, pages 1–8, 2011.

[9] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *International Conference on Computer Vision & Pattern Recognition (CVPR)*, pages 511–518, 2001.