16^{th} *Computer Vision Winter Workshop*
Andreas Wendel, Sabine Sternig, Martin Godec (eds.)
Mitterberg, Austria, February 2-4, 2011

Addressing false alarms and localization inaccuracy in traffic sign detection and recognition*

Igor Bonači
Faculty of electrical engineering and computing
Unska 3, Zagreb, Croatia
igor.bonaci@fer.hr

Ivan Kusalić
ivan.kusalic@fer.hr

Ivan Kovaček
ivan.kovacek@fer.hr

Zoran Kalafatić
zoran.kalafatic@fer.hr

Siniša Šegvić
sinisa.segvic@fer.hr

Abstract. *We present a study on applying Viola-Jones detection and SVM classification for recognizing traffic signs in video. Extensive experimentation has shown that this combination suffers from high incidence of false alarms and low tolerance to localization inaccuracy of the true positive detection responses. We report on three improvements which effectively alleviate these problems. Firstly, we confirm the previous result that raw detection performance of Viola-Jones detector can be improved by exploiting color. Additionally, we propose a solution for filtering false positive detection responses, based on a properly trained artificial neural network classifier in the last stage of the detection cascade. Finally, we propose a novel approach for alleviating the degradation of the classification performance due to localization inaccuracy. Experiments have been performed on several video sequences acquired from a moving vehicle, containing several hundred triangular warning signs. The results indicate a dramatic improvement in detection precision, as well as significant improvements in classification performance. At the system level, the proposed system correctly classified more than 97% of triangular warning signs, while producing only a few false alarms in more than 130000 image frames.*

1. Introduction

The ability to detect and classify objects is a key component of many computer vision applications. This paper considers a framework based on combining a boosted Haar cascade detection (the Viola-Jones algorithm) with support vector machine (SVM) classification. Although we address issues of general interest in object detection and recognition, our focus is on studying the considered framework in the context of ideogram-based traffic signs.

There are many exciting application fields of traffic sign recognition in video such as driving assistance systems, automated traffic inventories, and autonomus intelligent vehicles. These applications are important for the society since their main goal is to increase the traffic safety. Consequently, the challenges towards achieving human-like performance (e.g. illumination and color variance or motion blur) are actively researched. Recent high-class car models already come with optional traffic recognition systems, but only limited technical information about the employed algorithms and their performance is available. These recognition systems usually detect only speed limit signs and assume highway conditions, which significantly simplifes the problem.

In early experiments with the proposed framework we experienced two major problems: i) large number of false alarms, and ii) poor classification of the detection responses. This paper reports on several improvements which effectively alleviate these problems. We first report that color sensitive detection can reduce the false positive detection rate while improving the recall for large signs. The false detection rate is additionally reduced by a novel method consisting of adding an artificial neural network classifier as an additional level of a boosted Haar cascade. We present experiments which suggest that the poor classification performance is caused by the localization error in the detection responses. To solve this problem we propose an additional novelty, which is to modify the classifier training set according to the empirically determined properties of the localization error. The presented methods significantly improve the classification performance on standalone images, while the performance in video experiments approaches 100%

*This research has been jointly funded by the Croatian National Foundation for Science, Higher Education and Technological Development, and the Institute of Traffic and Communications under programme Partnership in Basic research, project number #04/20. The project web page is at http://www.zemris.fer.hr/ ssegvic/mastif

correct detection and classification.

2. Related work

Automated traffic sign detection and recognition has been an active problem for many years, and there is a vast number of related publications. The detection procedure solves the problem of locating traffic signs in input images, while the classification procedure determines the types of the detected traffic signs.

There are different approaches to detection. Some of the methods [6],[16] use color based segmentation, and model matching in order to detect the traffic sign. There are also researchers that rely only on the shape, using Hough transform [9][7], radial basis transform [11] etc. The other approach is to use a general purpose object detector. A popular algorithm for general object detection has been proposed by Viola and Jones [20]. The algorithm has been applied for traffic sign detection by several researchers [2, 18, 4]. A disadvantage of the original algorithm is that it disregards the color information, which might be valuable for detection. Bahlman et al. [1] use the Viola-Jones detector with extended feature set in order to use color information. That paper reports better detection performance using color, especially in reducing the false positive rate. This result encouraged us to use color information in Viola-Jones detector as well.

Munder and Gavrila [12] compared object detection methods performance on pedestrian classification. Their experiments show that the combination of Support Vector Machines with Local Receptive Field features performs best, while boosted Haar cascades can, however, reach quite competitive results, at a fraction of computational cost. We took advantage of both the Viola-Jones detector speed and the performance of a slower classifier by building a heterogeneous cascade of boosted Haar-like features with Artificial Neural Network as the final level of cascade. This approach significantly lowered the number of false detections.

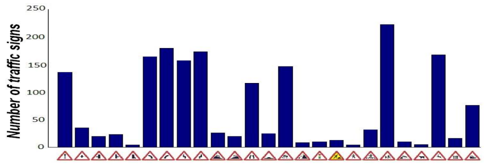
For the classification task, most of the previous approaches used one of well studied classification schemes, such as SVM [3], multiple discriminant analysis [17], neural networks [14] etc. A detailed report on current research in sign detection can be found in a recently published review paper by Nguwi and Kouzani [13].

3. The Dataset

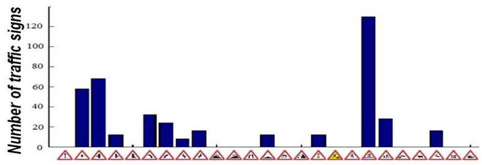
We used two datasets, labeled as dataset A and dataset B. The dataset A was used for learning and validation, while dataset B was used to test the performance. Both datasets were extracted from video sequences recorded with camera mounted on top of a moving vehicle. Video sequences were recorded at daytime, at different weather conditions. The dataset A corresponds to video material containing about 450 physical triangular warning signs, in which 1802 occurences have been manually annotated. The dataset B contains 265 physical triangular warning signs. Figure 1 shows examples of annotations.

Figure 1. Examples of extracted images

Traffic sign images were annotated manually in video sequences, while background images are extracted randomly from video sequences in dataset A. Altogether, 25 classes of traffic signs are represented in the dataset. Figures 2a and 2b show the distributions of the traffic sign classes present in datasets A and B.



(a)



(b)

Figure 2. Distribution of samples with respect to the sign class for dataset A (2a) and dataset B (2b).

4. Detection

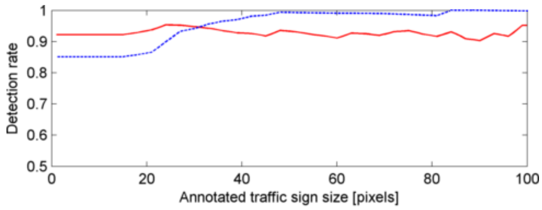
Our detection scheme is based on Viola and Jones' algorithm [20], a very popular method for real-time object detection.

In the next sections we will show the results of a standard Viola-Jones detector on our dataset and the modifications that were made to further improve the detection rate and the false positive rate.

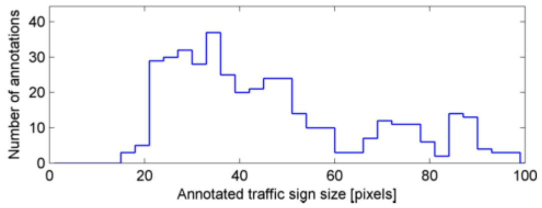
4.1. Viola and Jones' algorithm

Viola-Jones detector uses a cascade of boosted Haar-like features calculated on a gray-scale image.

For a human observer color is of great importance in traffic sign detection, so that by intuition we expect that color information should useful in machine detection as well. Bahlmann et al. [1] suggest computing the Haar-like features from multiple color channels (R, G, B, normalized r, g, b and grayscale).



(a)



(b)

Figure 3. Comparison of the Viola Jones detection with and without color information. 3a shows the detection rates with respect to the traffic sign size. The y-axis represents the detection rate for traffic sign that have an area larger than the value plotted on the x-axis. The dotted blue line corresponds to the color-based cascade, while the solid red line represents the grayscale cascade. 3b shows the distribution of the test dataset with respect to the traffic sign size.

We developed our own implementation of the algorithm which enables us to evaluate the impact of color information to the detection performance. The implementation employs the channels from the Lab color space, with which we obtained best results. Figure 3a compares the detection rates obtained by our color-enabled implementation and the corresponding grayscale version. The y-axis represents the detection rate for traffic signs that have an area larger than the value plotted on the x-axis[1]. The results show that color information has a positive impact when detecting larger traffic signs, but it has a negative impact when detecting small traffic signs (smaller than 30 pixels in size). The reason is that images of distant traffic signs are very small and contain very little color information, while larger images contain enough color (cf. Fig. 1).

In this work, we focus on the detection rate of larger traffic signs because our system will be used with video sequences and we expect that every traffic sign will become large enough for the system to detect. The problem with the system described so far is the false positive rate. When using the Lab cascade we get the false positive rate[2] of 68.7%, while with the grayscale cascade we obtain the false positive rate of 109.24%. Better detection rate for larger traffic signs and smaller false positive rate was the reason for choosing the color cascade. We still need to drastically reduce the false positive rate, because we use the system on video sequences.

4.2. Decreasing the false positive rate

In order to reduce the false positive rate we have added an additional stage to the detector cascade. The new stage is a binary classifier based on an artificial neural network[3]. The negative examples for ANN training have been collected as false positives of the Viola-Jones detector applied to the images from the learning dataset A. The positive training images are exactly the same as for the preceding stages of the cascade. The feature set for the neural network is based on the HOG (Histogram of Oriented Gradients) descriptor [5]. Figure 4 shows the arrangement of the HOG cells.

Figure 4. Arrangement of HOG cells in the detection window. The cell size is 6x6 pixels.

The Viola-Jones detector is used because it enables real-time detection, but in order to reduce the false positive rate it is better to use a heterogeneous cascade. Munder et al. [12]report that adding more stages to the VJ cascade further reduces the training set error, but the validation and test sets were observed to run into saturation. Using a stronger and less efficient classifier as the last stage of a VJ classifier does not have a negative impact on detection speed because only a small fraction of image patches passes the VJ cascade.

There are two possible ways of integrating ANN classifier with the Viola-Jones cascade. In the first arrangement the ANN is applied after the integration of multiple detections. That scheme drastically lowers the detection rate because of small errors in localization introduced by the integration of multiple detections. The neural network discards almost all traffic signs which are not aligned perfectly as the annotations used in the learning process. In the second arrangement the ANN is placed before the integration step, which proved to be much more effective. Usually there are several detections of a single traffic sign produced by the Viola-Jones detector, and some of these detections are perfectly aligned. Those detections are accepted by the ANN.

Figure 5 evaluates the impact of using the described combination. It is important to note that the detection rate is lowered only for traffic signs smaller

[1]Detailed results and parameters used are presented in the results sections.

[2]False positive rate is defined as the number of false detections divided by the number of existing traffic signs.

[3]SVM could be used instead of ANN as they yield almost identical results.

than 45×45 pixels. The false positive rate on a test dataset is reduced from 65.74% to 7.04%.

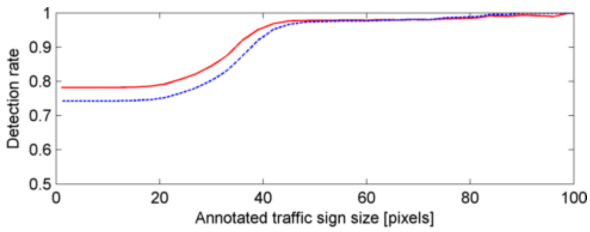


Figure 5. Detection rate with (solid red line) and without (dotted blue line) the ANN stage with respect to traffic sign size in pixels. The y-axis represents the detection rate for traffic signs that have an area larger than the value plotted on the x-axis.

Additionally, it is interesting to note that localization has improved after adding the additional level of cascade. We define localization as the percentage of overlap between an annotated traffic sign and the detection response. Figure 6 evaluates this impact, showing the distribution of traffic sign detections with respect to the localization error. We can see that the ANN stage removes some of the most inaccurately localized detection responses.

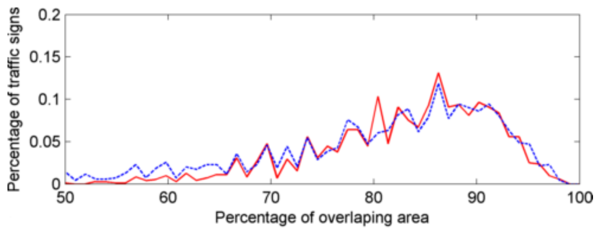


Figure 6. Localization error with (solid red line) and without (dashed blue line) adding the ANN stage. The x-axis represents the percentage of overlap between an annotated traffic sign and the detection response (localization quality).

5. Classification

When a traffic sign is detected, the next step is to determine the class it belongs to. In this section we describe the problems which arise due to localization inaccuracy of the detection responses and propose the solution.

5.1. Feature set and classifier

The first step in solving the classification problem is to choose which features to extract from the resized image patches corresponding to the detection responses. For that purpose we chose HOG descriptors [5] since they performed better than the raw pixels in early experiments with an ANN classifier. Before calculating the HOG descriptor the resized grayscale patches are first resized to 48× pixels, and then contrast-normalized and smoothed with the

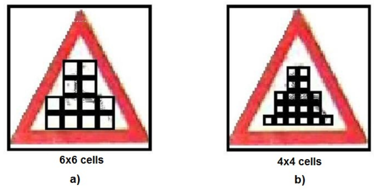


Figure 7. Arrangement of HOG cells over the detection window. Both sets of histograms are used for classification.

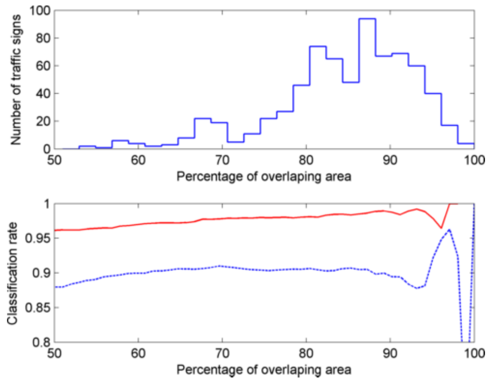


Figure 8. Classification performance of ANN (dotted blue line) and SVM (solid red line) with regard to the percentage of overlapping area between calculated area and annotation, after integration with the detection process. This graph represents the classification rate for all detections that have percentage of overlapping area with annotation larger than value plotted on the x-axis. Classification rate of the SVM classifier is consistently higher than the classification rate of the ANN. The decrease of classification rate at 98% overlap is a result of a single error in classification and therefore falls within the limits of a statistical error. The distribution of traffic signs (top image) is a coarsely discretized distribution from Fig. 6 (solid red line).

Gaussian filter. Figure 7 shows arrangement of HOG cells in a resized patch. Figure 7a shows cells of 6×6 pixels, while figure 7b shows cells of 4×4 pixels.

For each cell, a histogram of gradient orientations is calculated, and added to the feature vector. For the cells shown in figure 7a histograms have 4 bins and cover $(0, \pi)$ radians, while cells shown in figure 7b have histograms with 7 bins which cover $(0, 2\pi)$ radians. Both sets of cells shown in figure 7 are used in calculation of the feature vector. The dimension of the resulting feature vector is 174.

Having decided on the features that we will use, next we needed to choose a classifier, for which ANN and SVM were considered. After integration of both classifiers with the detection process, results shown in figure 8 were obtained. Dataset B was used as a test set, while the dataset A was used for learning (cf. Fig. 2). The figure clearly shows that SVM performs better then ANN, so that we chose SVM as our classifier.

Initial testing results showed that SVM with HOG

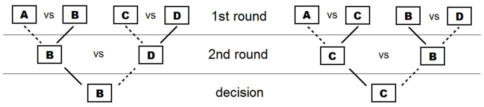


Figure 9. SVM as multi-class classifier. Two DAGSVM trees are shown, both of which use the same binary classifiers (A vs B, B wins; A vs C, C wins; A vs D, A wins; B vs C, C wins; B vs D, B wins; C vs D, D wins). To build DAGSVM tree, all available classes are divided into pairs. On the left side, pairs are: (A,B), (C,D) and on the right, pairs are (A,C) and (B,D). In each round, all pairs are evaluated using binary SVMs, and the winners advance to the next round (solid lines), which are in turn again divided into pairs. Losers are simply discarded (dashed lines). This process continues until only one class remains. Different arrangement of initial pairs can end up with different decisions, as is the case with the illustrated DAGSVM trees.

performs good enough without using kernels. There is no need to use kernels, because the results show that the classes are linearly separable in the feature space.

Because SVM is a binary classifier, we needed to decide on a strategy for using SVM as multi-class classifier. We decided against standard one-vs-one and one-vs-all methods, as they can both produce ambiguity errors in voting process if classes significantly vary in number of training examples. Instead, we used a method similar to the DAGSVM [15, 8]. Our method consists of building a directed acyclic graph which looks like upside-down binary tree with all available classes as leaves. In each step, all classes that are not eliminated are divided into pairs which are then used for one-vs-one binary classification. This way, after each step the number of classes considered is halved, until finally only one class remains. The remaining class is the classifier's decision. Each side of figure 9 illustrates this process.

Because different binary classifiers vary in reliability, this method can produce different results depending on the way classes are initially divided into pairs, as shown on figure 9, where two DAGSVMs make different decisions using the same binary classifiers. Obviously, one of those decisions is wrong, but it is not clear which one. That is why we construct this binary tree a few times (usually 5 times) and employ a simple voting strategy. Each time different separation into pairs is used. Different pairing distributions had little effect, as most of binary SVMs are quite reliable. Nevertheless it did improve classification rate a little, and had no trade offs, as it only consumes slightly more time, which is not of concern in the classification process.

5.2. Modelling the localization error

Aside from relative performances of ANN and SVM, figure 8 shows another interesting phenomena, namely that both classifiers have lower classification rates then we first anticipated. This was at first confusing, as both classifiers performed much better in initial tests that were used to verify validity of implementations, with classification rates around 95% (ANN) and 98% (SVM). We realised that the problem was caused by the localization inaccuracy of the detection responses. Many detections have a small offset, mostly only a pixel or two in each direction. Fig. 10 shows localization error of the detections, while Fig. 11 shows relative scale deviation[4], both with regard to the groundtruth annotations. The presented data was obtained by evaluating the previously described detector on images from dataset A and comparing the detections with annotated locations of traffic signs.

To solve this localization problem, we decided to expand training set with examples that resemble detector's errors, with traffic signs annotated slightly off. Specifically, for each annotated example in training set, we added another 10 examples which model detector's errors. As it can be seen from figures 10 and 11, both types of errors can be modeled with normal distribution. Localization error (expressed relative to the vertical or horizontal sizes of traffic signs) was modeled as normal distribution with parameters ($\mu = -0.014, \sigma = 0.0016$) for x-axis and ($\mu = -0.026, \sigma = 0.002$) for y-axis. Relative scale deviation was modeled as normal distribution, with parameters ($\mu = 1.065, \sigma = 0.074$). The distributions shown on figures 10 and 11 were obtained by comparing detection responses to the annotations from the training set. The SVM classifier trained on the modified training set got the correct classification rate of 95.42%, as opposed to 91.33% obtained with the unmodified training set. Figure 12 shows detailed comparison of results achieved with SVMs trained on different training sets.

The idea of increasing the training set by adding translational jitter has been proposed before, but with different purpose and motivation. For example, Laptev [10] employs this idea to enlarge the training dataset for learning an object detector, while our primary motivation is to improve the recognition performance in presence of localization inaccuracy of the detector responses.

6. System overview

After detection and classification is conducted on images, the next step is to identify the traffic sign through the consecutive images (i.e. video). An output of the Viola-Jones detection is considered a false positive, and thereby is discarded, if a detection is not present in at least three consecutive frames. Group-

[4]Relative scale deviation describes the ratio between detected size and the annotated size.

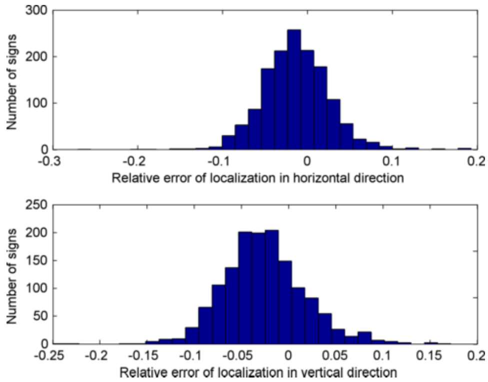


Figure 10. Relative translational deviation of the detection responses with regard to annotation.

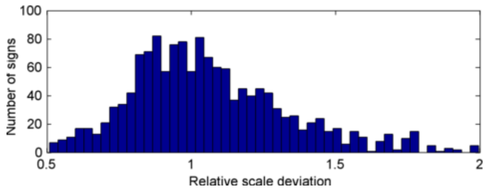


Figure 11. Relative scale deviation of the detection responses with regard to annotation.

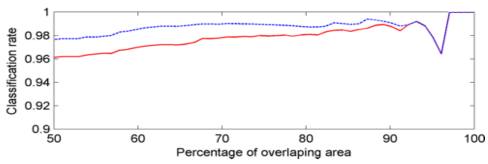


Figure 12. Comparison of classification results achieved with SVMs trained on unmodified (solid red line) and modified (dotted blue line) training sets. This graph represents the classification rate for all detections that have percentage of overlapping area with annotation larger than value plotted on the x-axis. The decrease of classification rate at about 97% overlap is a result of a single error in classification and therefore falls within limits of a statistical error.

ing of single frame detections to a joint detection through the video is based on thresholding the overlapping area between two consecutive detections. For each detection a classification phase is conducted and the final class is determined by voting with equal weights for all detections.

The final process consists of 5 phases:

1. Traffic sign detection with Viola-Jones detector
2. Filtering false detections with ANN classifier
3. Integration of multiple detections
4. Traffic sign classification (SVM)
5. Identifying traffic signs in video

Example of system behaviour is shown in figure 15.

Final system has a frame rate of 15 frames per second with an input image size of 480x360 pixels on Intel 1.8 GHz Dual Core computer. The detection process is implemented to take advantage of multi-threading features of a processor.

7. Experimental results

The performed experiments are divided in two categories: results on standalone images and results on video. All experiments were conducted on the same test dataset B corresponding to about 1.5 hours of video material. Detailed information about the test dataset is as follows:

- duration: 1 hour, 28 minutes and 16 seconds
- resolution: 480x360 pixels
- frame rate: 25 fps
- number of frames: 132420
- number of physical traffic signs: 265

For each frame from the video sequence, position of all the traffic signs is given, along with the annotated classification.

7.1. Results on standalone images

We provide results achieved on standalone images first, as the employed core algorithms naturally take an image on input.

Parameters for detection algorithm are as follows:

- Viola-Jones scale factor: 1.2
- Viola-Jones sliding window step size: 5% of current window size
- minimal number of detections needed for confirming the detection: 3

Figure 13 shows achieved detection rates with regard to size of annotation. Total detection rate is 83.53%, which does not look all that impressive at first. Main reason for such low detection rate is the fact that our Viola-Jones implementation uses sliding window with minimal size of 24×24 pixels. If the images with smaller signs are excluded from the test dataset the detection rate increases to 89.18%, which is still too low for practical usage. However, almost all signs larger than 50×50 pixels were successfully detected (99.14%), which gives hope that detections on video would be good enough. The reason for this optimism lies in the fact that the size of a traffic sign increases as video progresses and the vehicle advances closer to the sign.

In experiments in this subsection, the classification was evaluated only on successful detections. Figure 14 shows comparison of SVMs trained on unmodified (dotted blue line) and modified (solid red line) training set (extracted from dataset A). Similarly to the detection results, the total classification rate of SVM trained on modified set is 93.59%, as opposed to 85.14% for SVM trained on unmodified set. It is important to note the importance of dataset modeling according to the localization error.

7.2. Results on video sequence

The results presented in the previous section can be extended to take advantage of multiple occurences

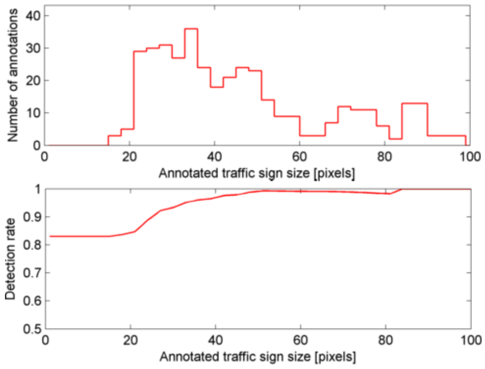


Figure 13. Detection rates with regard to the annotation size in the test dataset. The y-axis represents the detection rate for traffic signs that have an area larger than the value plotted on the x-axis.

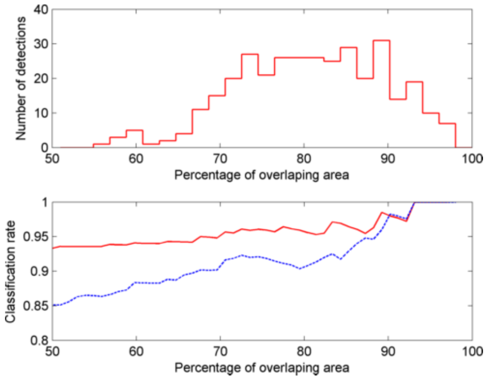


Figure 14. Classification rates with regard to percentage of overlap area between detection and annotation are shown. Solid red line represents classification rate for SVM trained on a modified set, and the dotted blue line represents classification rate for SVM trained on an unmodified set. Graphs represent the classification rate for all detections for which the percentage of overlapping area with the corresponding annotation is larger than value plotted on the x-axis.

Table 1. Final detection results on video sequence.

No. of traffic signs	265
No. of detected signs	260
No. of false detections	2
Detection rate	98.11%
False positive rate	0.75%
False positives per frame	0.0015%

of physical traffic signs in video. The detection results on the test video sequence are given in Table 1[5]. The achieved results are very good, since only smaller traffic signs of poor quality are not detected. This suggests that the detection rate could be increased if the video of higher resolution was used.

There are 243 physical traffic signs that are considered for classification as opposed to 265 total traffic signs present in the video sequence. The 22 missing signs belong to classes for which the classifier has not been trained because dataset A used for training purposes contains insufficient number of training examples from those classes (or none at all). Final results for classification on video sequence are given in table 2. Only two traffic signs were misclassified, both of which are similar to their respective target classes. By combining detection and classification processes, we get overall system performance of 97.3%. Typical results at the system level are illustrated in Fig. 15.

Table 2. Final classification results on video sequence.

No. of traffic signs	243
No. of correct classifications	241
Classification rate	99.17%

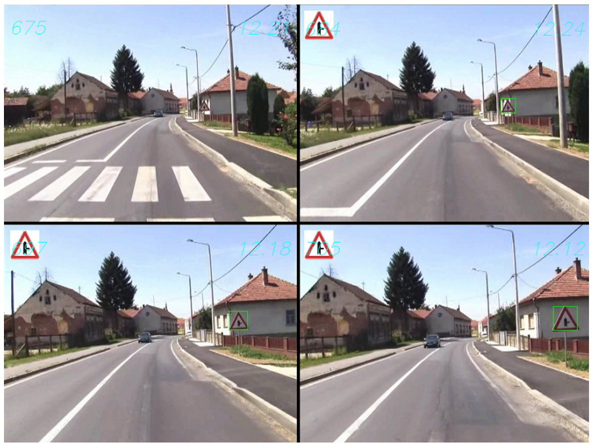


Figure 15. Typical behavior at the system level. In the beginning, the traffic sign is too small, but as time goes on, it becomes big enough and gets detected.

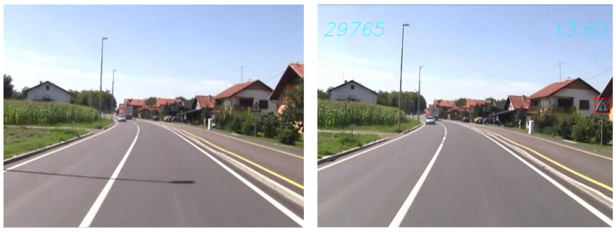


Figure 16. A misdetection of a traffic sign. Size of the traffic sign on the right is 25×25 pixels. Four out of five misdetections at the system level are very similar to this one.

Fig. 16 is an example of a misdetection. The traffic sign in question does not get bigger than 25×25 pixels so that it is detected only in a single frame and is consequently discarded.

8. Conclusion and future work

This paper presents two novel methods that can improve performance of detection and classification either in standalone images or in video. The first

[5]False positive rate on video sequence is defined as the number of false detected traffic signs divided by the total number of traffic signs.

method is used to decrease the false positive detection rate by extending the boosted Haar cascade with an additional stage containing a strong nonlinear binary classifier. The second method adapts the training set to the empirically estimated model of the localization error in the detection responses. Both methods were applied in the frame of a real-time system working on video sequences. The obtained results strongly suggest that automated road inspection is likely to become feasible in the near future.

Some categories of the triangular warning signs are represented with less than 10 samples in the employed training dataset. Thus we believe that it would be possible to obtain even better classification results by collecting a more complete training dataset.

The implemented method for combining information from consecutive frames is very simple, and could be improved in several ways. One of the directions we are currently pursuing is to obtain better detection for small signs by making the ANN detection filter less strict and resolve the remaining false positives with additional approaches. These additional methods would be based either on spatio-temporal properties of the recorded trajectories of the traffic signs, or on enforcing the temporal consistency of the detection responses corresponding to the common physical traffic sign.

We are also interested in expanding the scope of this research to other traffic sign types, such as round traffic signs (mandatory signs). Traffic signs of type C (informational signs) could prove to be especially challenging, as they come in many different shapes and sizes. Using a Viola and Jones' detector for each type of a traffic sign would slow the system considerably. Torralba et al. [19] proposed a method for multiclass object detection, which could be of use in dealing with this diversity without adding much overhead to detection time.

Finally, we are also interested in detecting the state of deterioration of the detected traffic signs. The issues we would like to deal with are fading colors, deformation of the sign pole or the sign itself, inappropriate orientation, or partial occlusion.

References

[1] 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. *Proc. of IEEE IV*, 2005. 2

[2] X. Baro and J. Vitria. Fast traffic sign detection on greyscale images. *Recent Advances in Artificial Intelligence Research and Development*, pages 131–138, October 2005. 2

[3] S. M. Bascon, J. A. Rodriguez, S. L. Arroyo, A. F. Caballero, and F. Lopez-Ferreras. An optimization on pictogram identification for the road-sign recognition task using SVMs. *Computer Vision and Image Understanding*, 2010. 2

[4] K. Brkić, A. Pinz, and S. Šegvić. Traffic sign detection as a component of an automated traffic infrastructure inventory system. In *Proc. of OeAGM/AAPR*, pages 129–140, 2009. 2

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. of CVPR*, pages 886–893, 2005. 3, 4

[6] 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. *J. Vis. Commun. Image R.*, pages 675–685, 2006. 2

[7] M. Garcia-Garrido, M. Sotelo, and E. Martin-Gorostiza. Fast traffic sign detection and recognition under changing lighting conditions. In *Proc. of IEEE ITSC*, pages 811–816, 2006. 2

[8] C.-W. Hsu and C.-J. Lin. A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, 13(2):415–425, 2002. 5

[9] W.-J. Kuo and C.-C. Lin. Two-stage road sign detection and recognition. In *Proc. od IEEE ICME*, pages 1427–1430, 2007. 2

[10] I. Laptev. Improving object detection with boosted histograms. *Image Vision Comput.*, 27(5):535–544, 2009. 5

[11] G. Loy and A. Zelinsky. A fast radial symmetry transform for detecting points of interest. In *Proc. of ECCV*, page 358, 2002. 2

[12] S. Munder and D. Gavrila. An experimental study on pedestrian classification. *IEEE Transactions on PAMI*, 28(11):1863–1868, 2006. 2, 3

[13] Y. Nguwi and A. Kouzani. A study on automatic recognition of road signs. In *Proc. IEEE CIS*, pages 1–6, 2006. 2

[14] H. Ohara, I. Nishikawa, S. Miki, and N. Yabuki. Detection and recognition of road signs using simple layered neural networks. In *Proc. of ICONIP vol. 2*, pages 626–630, 2002. 2

[15] J. C. Platt, N. Cristianini, and J. Shawe-Taylor. Large margin dags for multiclass classification. In *Advances in Neural Information Processing Systems*, pages 547–553. MIT Press, 2000. 5

[16] A. Ruta, Y. Li, and X. Liu. Real-time traffic sign recognition from video by class-specific discriminative features. *Pattern Recognition*, 2010. 2

[17] S. Segvić, K. Brkić, Z. Kalafatić, V. Stanisavljević, M. Ševrović, D. Budimir, and I. Dadić. A computer vision assisted geoinformation inventory for traffic infrastructure. In *Proc. of ITSC*, Portugal, 2010. 2

[18] R. Timofte, K. Zimmermann, and L. Van Gool. Multi-view traffic sign detection, recognition, and 3d localisation. In *Proc. of WACV*, pages 69–76, Snowbird, Utah, 2009. 2

[19] A. Torralba, K. P. Murphy, and W. T. Freeman. Sharing visual features for multiclass and multiview object detection. *IEEE Transactions on PAMI*, 29(5):854–869, 2007. 8

[20] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004. 2