COLOR EXPLOITATION IN HOG-BASED TRAFFIC SIGN DETECTION

I.M. Creusen[1,3]*, R.G.J. Wijnhoven*[2,3]*, E. Herbschleb*[3]*, P.H.N. de With*[1,3]

[1]CycloMedia BV
P.O. Box 68, 4180 BB
Waardenburg, NL

[2]ViNotion BV
P.O. Box 2346, 5600 CH
Eindhoven, NL

[3]Eindhoven University of Technology
P.O. Box 513, 5600 MB
Eindhoven, NL

ABSTRACT

We study traffic sign detection on a challenging large-scale real-world dataset of panoramic images. The core processing is based on the Histogram of Oriented Gradients (HOG) algorithm which is extended by incorporating color information in the feature vector. The choice of the color space has a large influence on the performance, where we have found that the CIELab and YCbCr color spaces give the best results. The use of color significantly improves the detection performance. We compare the performance of a specific and HOG algorithm, and show that HOG outperforms the specific algorithm by up to tens of percents in most cases. In addition, we propose a new iterative SVM training paradigm to deal with the large variation in background appearance. This reduces memory consumption and increases utilization of background information.

Index Terms— Object detection, Object recognition

1. INTRODUCTION

In this paper we consider traffic sign detection on images obtained from a driving vehicle in large-scale environments. Panoramic images are captured using two cameras with fish-eye lenses, thereby creating lens-distortion. As a pre-processing step, image enhancement algorithms are used to improve the image quality. We study the automated detection of a subset of the traffic signs in The Netherlands.

An object detection system can be realized in two ways. A first approach is to manually create a *specific object model* by using prior knowledge regarding the objects and the scene. A second option is to learn a *generic object model* automatically from manually annotated training data, also called supervised learning. An advantage of a specific object model is that the prior knowledge is explicitly modeled and no annotated training samples are required. Optimizing such a model requires much effort and the resulting model cannot be reused for other types of objects. However, because all prior knowledge is used in the model, an effective and efficient detection algorithm can be obtained. On the other hand, a generic object model is learned automatically from training data with little manual interaction. However, the model uses no prior information, and extracts its knowledge from the training data only. Therefore, good quality and completeness of the training data are key requirements. A clear advantage of such a model is that it can be generated for different object classes without manual tuning. In this paper we compare these two approaches for a large-scale traffic sign detection application.

Many algorithms for traffic sign detection are primarily focusing on color only. A survey on traffic sign detection was published by Fleyeh and Dougherty [1].

The work of Viola and Jones [2] describes one of the first successful applications of the generic approach. They designed a face detection algorithm for gray-scale images, using a 24×24 pixel detection window. From this window they extract a number of Haar-like features. These features are a local approximation to the image gradient which can be efficiently computed using integral images. The face detector is trained using many example face images, but can easily be trained to detect other objects.

Another highly successful sliding-window object detector is the Histogram-of-Gradient detector (HOG) proposed by Dalal and Triggs [3]. The algorithm outperforms the Viola-Jones algorithm. The authors propose to divide the detection window into cells, and for each cell a histogram of the image gradient (over orientation) is made. This type of feature is a dense version of the popular SIFT [4] technique. However, HOG features are not rotation and scale invariant. After the feature generation stage, a Support Vector Machine (SVM) is used to classify the high-dimensional features. In a recent evaluation for pedestrian detection, the HOG algorithm gives competitive performance [5].

Another sliding-window detector is recently proposed by Baró and Vitrià [6], which has also been applied for traffic sign detection. They use a more general version of the Haar-like features, called Dissociated Dipoles. During the training process, a genetic algorithm is used to iteratively add new features to the system, in contrast to the exhaustive search done by Viola and Jones. This approach leads to strongly improved performance compared to the standard Viola-Jones approach, the AUC (Area Under Curve) for a traffic sign detection problem is increased from 60% to 90%.

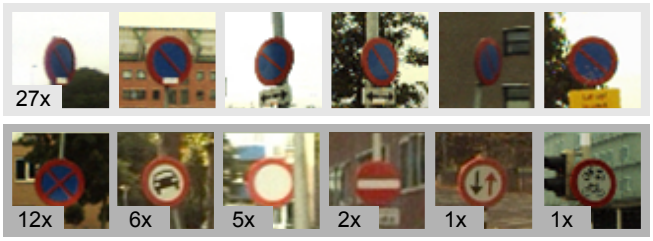
The Implicit Shape Model is proposed by Leibe and Schiele in [7]. Their idea is to locate small parts of the object in the image, and vote for a consistent center location of the total object. The maxima in this voting space defines the location of the object. This technique gives competitive results for generic object detection with relatively large objects.

In this paper our aim is to study a more generic algorithm that can handle the large variety of traffic signs. We have adopted the HOG algorithm because of its high performance, parallel implementation possibilities, and fast training compared to the proposal by Baró and Vitrià [6]. A closer look to the various algorithms reveals that the HOG algorithm implicitly exploits features like the gradient patterns and the shape of traffic signs instead of explicitly building models of those features as in done in the above proposals from literature. This is why we expect at least similar results.

Furthermore, we propose to extend the standard HOG algorithm by utilizing color information as a concatenation of per-channel HOG features. We show that the choice of the color space significantly influences the performance, and the optimal choice depends on the type of traffic sign.

In this paper, we compare a state-of-the-art specific algorithm and our generic HOG-based algorithm. The specific detection algorithm by Herbschleb and De With [8] uses a fast three-stage ap-

(a) Circular blue signs



(b) Circular red signs

(c) Triangular signs

Fig. 1. *Examples of traffic sign training samples. The numbers represent the number of samples in the training set.*

proach and uses color and shape features. Firstly, a fast algorithm discards uninteresting image regions by distinguishing specific color information. Secondly, traffic signs are detected by locating several consistent parts. Finally, a comprehensive validation step ensures the validity of the detected signs. In contrast with this, our algorithm models the background by a new iterative learning procedure. As our algorithm has a generic nature, it can be reused for other objects than traffic signs.

2. ALGORITHMS

The original HOG algorithm by Dalal and Triggs [3] applies a dense description strongly based on the popular *SIFT* algorithm from Lowe [4]. In a preprocessing step, the image input pixels are converted into HOG features, and object detection is performed by sliding a detection window over the image. To obtain scale-invariant detection, the preprocessing and detection process is repeated for downscaled versions of the input image. After applying the detector, detections at various scales are merged using a mean-shift mode finding algorithm [9].

Let us now discuss the HOG feature generation step in more detail. The processing steps are shown in Figure 2. For each input pixel, the gradient magnitude and orientation are calculated. The gradient magnitude of the pixel is added to the corresponding orientation bin. Input pixels are spatially quantized in cells of $n \times n$ pixels, where n is the cell size. Each cell contains one orientation histogram. To avoid quantization effects, linear interpolation is applied, both in the orientation and in the two spatial dimensions. Illumination invariance is obtained by using the gradient operator. Contrast normalization is applied by normalization of the orientation histograms. Dalal and Triggs propose to normalize each cell several times, for each $b \times b$ local neighborhood, where typically $b = 2$. The total feature vector of a detection window is the concatenation of the normalized orientation histograms of all the cells in the window.

For learning the actual detector, we use a linear Support Vector Machine (SVM). Although kernel SVMs will increase performance (as shown e.g. in [3]), a linear SVM is used for computational efficiency, as we execute our algorithm on a large-scale database. We use the same implementation used by Dalal and Triggs. The SVM classifier is trained in an iterative process, unlike the proposal of Dalal and Triggs. In the first iteration, all positive images are processed and a set of randomly chosen background regions is used as negative samples. In each additional iteration, the current detector is applied to a new image without traffic signs and resulting false detections are added to the training set for the next iteration. After each iteration, the classifier is retrained and all negative training samples which are not support vectors are discarded. The consequence of this technique is that the set of negative features remains small. This has two advantages: (1) it avoids the memory limitations observed by Dalal and Triggs and (2) it allows the utilization of more background samples, leading to a more accurate description of the background training set.

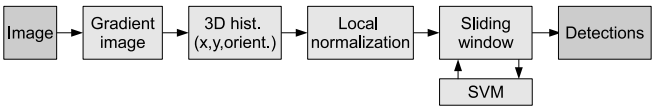


Fig. 2. *Overview of the HOG algorithm.*

As previously mentioned, we extend the standard algorithm by including color information in the HOG-descriptor. This is done by concatenating HOG descriptors for each color channel of the used color space. Note that the dimensionality of the resulting features and the computational complexity for detection increases linearly with the number of color channels.

The specific detection algorithm by Herbschleb and De With [8] uses a three-stage approach. Firstly, a fast algorithm discards uninteresting image regions by using color information. The first step applies color quantization and classifies each pixel of the image in the color classes red, blue, yellow, white and black, as they are the most appearing colors in traffic signs. This step uses a fixed mapping which was empirically estimated. Secondly, SIFT features [4] are extracted at Hessian interest points, and these are matched to a dictionary of traffic-sign features. The dictionary is constructed from synthetic traffic-sign images as specified by the local authorities. The spatial consistency of neighboring features is checked to improve robustness. If three or more matches indicate the same traffic sign, it is added as a valid detection. The final stage performs a validation of the generated detections by checking color consistency and template matching with several distorted templates.

There is a fundamental difference between the two approaches. The generic HOG detector detects traffic signs of a specific category only, whereas the specific algorithm is designed to detect all variations of several traffic signs in a single algorithmic pass.

3. EXPERIMENTAL RESULTS

Targeting the application of traffic sign detection on country-wide scale, we evaluate our algorithms with a set of about 5,000 annotated high-resolution images taken from a driving vehicle, using two fish-eye lens cameras. For the experiments, we extract traffic signs from a

Fig. 3. *Example image showing several correctly detected traffic signs, indicated by cyan rectangles and one undetected sign indicated by a red rectangle*

subset of the images to obtain a training set, and use the other images (approx. 3,000) for testing.

We evaluate the detection of three different classes of traffic signs: *blue-circular*, *red-circular* and *triangular* signs, and we train the system using 170, 74 and 53 samples, respectively. Each class contains intra-class variation due to the various signs in the class and due to the different imaging conditions, as shown in Figure 1. The distribution of the different types of signs in the training sets is representative for the total dataset. The resolution of the images in our dataset is 4800×2400 pixels.

Using the generic HOG detection algorithm, we train a different detector for each class from the positive object samples and a common set of negative samples in the form of images containing no traffic signs. Additionally, for each class, the positive samples of the other classes are added as negative examples. The positive samples are traffic signs having a resolution of 24×24 pixels in a 48×48 pixel region. Dalal and Triggs [3] have found that the use of contextual information is beneficial. For training the SVM, the proposed iterative approach is used with an initial set of 200 randomly selected background samples.

We consider different versions of the HOG algorithm. Whereas Dalal and Triggs propose to use the gradient in the color channel with maximum gradient magnitude, traditional HOG only uses a single color channel. The green channel of the RGB color space is often employed for traffic sign detection, but this causes many misdetections between red and blue signs. We have found that the H-channel of HSV color space gives better results. In our experiments we will use the H-channel detector as a single channel detector. Results for the G-channel detector are omitted because the performance is significantly less. To incorporate more color information, we concatenate HOG descriptors for each color channel. We compare results for the following color spaces: RGB, HSV, CIELab and YCbCr.

In our experiments, we have used the following settings for our HOG detector: cell size 4×4 pixels, 9 orientation bins and 4 block normalizations ($b = 2$). For each color channel, the dimensionality of the feature vector is $2,304$. Applying the single-channel HOG detector on a 4800×700 image takes about 23 seconds using a single CPU-core at 2.7 GHz. Each image is downscaled in 35 steps using a scaling factor of 1.05. This leads to the detection of traffic signs ranging from 24×24 pixels to 132×132 pixels. Because of the preprocessing steps in the specific algorithm, the execution time varies significantly over the total set of images. Typical execution times vary between 30 seconds and a few minutes. Note that in a single pass of the specific algorithm, all traffic sign classes are detected simultaneously, whereas the generic detector locates only a single class of signs.

We have applied both the specific algorithm and the HOG detectors to the dataset (see Figure 1) and the results are shown in Figure 4. The AUC scores are summarized in Table 1. Figure 2 shows an example output image of the CIELab detector. In general, we observe that the HOG detector outperforms the dedicated algorithm in most cases. We have found that the choice for a color space has a significant impact on the detection performance. For blue circular traffic signs, the performance in the CIELab color space is superior to other color spaces. For red circular traffic signs, the CIELab and YCbCr color spaces give similar performance, while for red triangular signs the performance in the YCbCr space is the highest. Detection in the RGB and HSV color spaces is suboptimal in these experiments. It is interesting to note that performance in the H-channel is almost identical to the performance in the HSV-space. This indicates that saturation and intensity information is largely irrelevant for the considered traffic sign detection application.

Name	Red circ. AUC	Blue circ. AUC	Triangular AUC
Dedicated	41.6%	56.2%	45.5%
H(HSV)	32.0%	70.3%	50.0%
HSV	32.0%	70.4%	50.0%
CIELab	**56.0%**	**85.0%**	65.7%
RGB	46.4%	56.9%	52.8%
YCbCr	55.7%	69.2%	**74.6%**

Table 1. *Detection performance of the Dedicated algorithm and the HOG detector in several color spaces, for three different classes of traffic signs. The highest scores are indicated in bold.*

4. CONCLUSIONS

We have evaluated two different algorithms for traffic sign detection on a large-scale dataset. A dedicated algorithm uses a processing chain of three stages to detect traffic signs, which has been manually tuned. We compare this to the generic Histogram of Oriented

Gradients (HOG) algorithm, that automatically learns its detector from a set of training images. In addition to the standard HOG algorithm, we propose an extension that simultaneously uses information of multiple color channels and show that it outperforms the single-channel algorithm. Furthermore, we have employed an iterative technique for SVM training which is novel in this context, to deal with the large variation in background appearance. This significantly lowers memory consumption and therefore allows the utilization of more background images in the training process.

Experimental results show that for the considered task, the generic HOG algorithm significantly outperforms the dedicated algorithm in most cases by a range of 10–30%. The choice of the color space has a profound effect on the performance. We have found that the CIELab and YCbCr spaces provide the best performance, probably due to the availability of two dedicated color channels fitting to the traffic signs. The HSV and RGB spaces are less suitable for traffic sign detection. Furthermore, we have shown that performance of the single channel H-detector is nearly identical to the performance of the HSV-detector. This indicates that saturation and intensity information is largely irrelevant for the considered traffic sign detection application and thus that color is the dominant feature.

5. REFERENCES

[1] H. Fleyeh and M. Dougherty, "Road and traffic sign detection and recognition," in *Proc. 16th Mini EURO Conf. and 10th Meeting of EWGT*, September 2005, pp. 644–653.

[2] Paul Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2001, vol. 1, pp. 511–518.

[3] N. Dalal and B. Triggs, "Histogram of oriented gradients for human detection," in *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, June 2005, vol. 1, pp. 886–893.

[4] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. Journal of Computer Vision (IJCV)*, vol. 60, no. 2, January 2004.

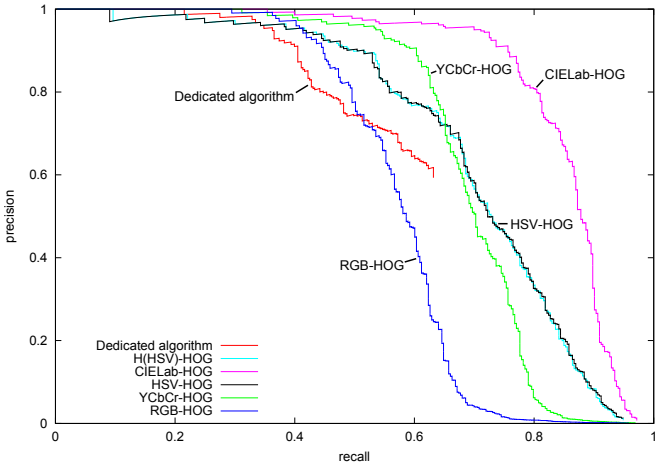
[5] Piotr Dollar, Christian Wojek, Bernt Schiele, and Pietro Perona, "Pedestrian detection: A benchmark," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2009, pp. 304–311.

[6] Xavier Baró and Jordi Vitrià, "Probabilistic darwin machines for object detection," in *Proc. Int. Conf. on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, July 2009, vol. 2, pp. 839–845.

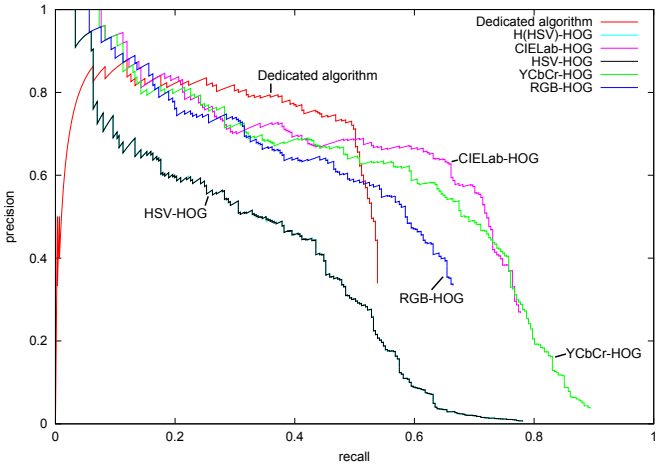
[7] Bastian Leibe and Bernt Schiele, "Interleaved object categorization and segmentation," in *Proc. British Machine Vision Conference (BMVC)*, September 2003, pp. 759–768.

[8] Ernst Herbschleb and Peter H.N. de With, "Real-time traffic-sign detection and recognition," in *Visual Communications and Image Processing, SPIE-IS&T Electronic Imaging*, January 2009, vol. 7257, pp. 0A1–0A12.

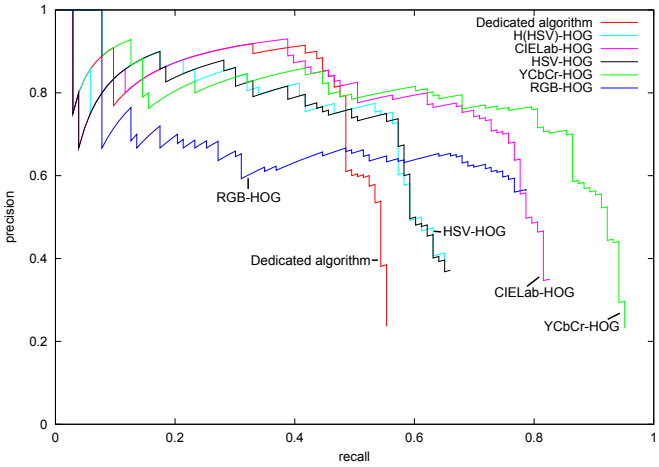
[9] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 24, no. 5, pp. 603–619, May 2004.



(a) Circular blue traffic signs.



(b) Circular red traffic signs.



(c) Triangular traffic signs.

Fig. 4. *Resulting recall-precision curves for the evaluated traffic sign classes. Note that the H(HSV) results show significant overlap with the HSV results. Figure best viewed in color.*