

Video-based Detection and Classification of US Traffic Signs and Mile Markers using Color Candidate Extraction and Feature-based Recognition

Vahid Balali¹ and Mani Golparvar-Fard²

¹PhD Candidate, Department of Civil and Environmental Eng., Univ. of Illinois at Urbana-Champaign, 205 N Mathews Ave., Urbana, IL 61801; PH (540) 235-6474; FAX (217) 265-8039; email: balali2@illinois.edu

²Assistant Professor and NCSA Faculty Fellow, Depts. of Civil and Environmental Eng. and Computer Science, Univ. of Illinois at Urbana-Champaign, 205 N Mathews Ave., Urbana, IL 61801; PH (217) 300-5226; email: mgolpar@illinois.edu

ABSTRACT

Traffic sign and mile-marker detection and classification are among the important components of highway asset management systems. The significant number of these high-quantity and low-cost assets in US highways can negatively impact on quality of any manual data collection and analysis. To address these challenges, this paper presents an efficient pipeline for video-based detection and classification of traffic signs and mile-markers based on color and shape criteria. Candidate extraction is based on finding the optimum RGB thresholds which yields high detection rate (very low False-Negatives), while keeping the number of False-Positives in check. The connected components from a thresholded image are extracted next. We use sliding windows, Haar-like features, and AdaBoost learning method to classify the detected assets. Experimental results with an average classification accuracy of 79.30% on actual data collected from US-460 highway show the promise of the proposed method for reducing the time and effort required for developing traffic road asset inventories.

INTRODUCTION

The US Departments of Transportation (DOTs) require reliable and up-to-date information about location and condition of highway traffic signs and mile markers for inventory and condition assessment purposes. Centralized and integrated with economic asset data, such databases can help DOTs to prioritize different highway sections for maintenance and replacement planning purposes (Cheok et al. 2010).

The data collection and analysis has to be done for millions of kilometers of roads and should be repeated periodically to identify missing or damaged signs to minimize safety threats. The advent of high resolution cameras and the increase in computational power have enabled DOTs to leverage video-based methods for data collection. Nonetheless, the analysis – detecting and localizing traffic signs – has primarily remained a manual process. Automating the analysis process is challenging. As shown in the US Manual on Uniform Traffic Control Devices ((FHWA) 2003), traffic signs are fabricated with large variety in appearance including materials, shapes, sizes, legends, and colors (Tsai et al. 2009). The in-class variability and between-class similarity of these signs can significantly challenge detection and

classification methods. In addition, signs vary from region to region and from country to country.

Over the past few years, the potential of applying object detection algorithms has enabled development of unique computer vision methods for detection and classification of traffic signs. Consequently, several high-end vehicles are already equipped with driver assistance systems which offer automated detection and classification for a few classes of traffic signs (Brkic 2013). However, it would be a blunder to assume the process is not challenging (Balali et al. 2013). To make these methods useful, both false positive (FP) and false negative (FN) rates should be very low. That is a major reason why current traffic sign inventory and condition assessment practices are still carried out manually.

Beyond the challenging appearances of the traffic signs, there are also several practical problems associated with variations in lighting condition, background, pose, viewpoint, and occlusions caused by other objects. Figure 1 shows examples of these challenges associated with a computer vision based solution for traffic sign detection.



Figure 1. Variability of traffic signs with different illumination, rotation, occlusion, and viewpoint.

To address these limitations, this paper presents an efficient pipeline for the detection and classification of US traffic signs. To be succinct, in this paper, we use “traffic signs” to represent both traffic signs and mile marker categories. A comprehensive data set of different types of traffic signs has been collected and created for benchmarking and validating the performance of the proposed detection and classification pipeline. In the following sections after a brief review of relevant literature, our method is presented and experimental results are discussed.

RESEARCH BACKGROUND

There are more than 670 types of traffic signs which come in hundreds of variations, such as in dimension, color, text, and font. (Hu and Tsai 2011) present a sign recognition algorithm that primarily focuses on speed limit signs. Their algorithm benefits from a probabilistic color model, the likelihood of sign locations in the images along with the traditional sign features (shape, color, and content features). Others focus on features such as color (Maldonado-Bascon et al. 2007), shape (Kim et al. 2005), combined color and shape features (Zhu et al. 2005), or geometrical and physical features along with text (Liu et al. 2006). In several cases such as (Fatmehsan et al. 2010; Wu and Tsai 2006), the focus is on detection of a single type of sign such as speed limit sign or stop signs. (Brkic 2013) review the

popular traffic sign detection methods in three categories: color-based, shape-based, and learning-based. (Timofte et al. 2011) propose a method for 2D recognition and 3D localization of traffic signs. It focuses on 3D sparse point cloud reconstructions and recognition of traffic signs, and high average performance are reported on 2D recognition. Since their method primarily focuses on 2D recognition, it is not directly applicable for condition assessment of certain assets such as guardrail and light poles. In the context of infrastructure projects, (Balali and Golparvar-Fard 2014; Golparvar-Fard et al. 2012) propose highway asset detection and recognition algorithm based on sematic texton forest and scalable non-parametric parsing that can simultaneously segment an image and detect 3D highway assets such as guardrails. Compared to prior work, our contribution in this paper is a new computationally efficient method that can detect and classify large variety of US traffic signs and mile markers from video streams.

METHOD

Our method involves 2D candidate extraction, followed by detection and classification. Using the video streams collected from the cameras mounted on the vehicle, a set of *thresholded* frames is initially identified. Each thresholded frame contains a set of candidates for traffic signs along the right side of the highway. Using a Boolean linear optimization based on the RGB color channel at pixel level, a set of bounding boxes are initially extracted wherein each bounding box potentially includes a potential traffic sign. This stage still recognizes all traffic signs that are partially occluded, damaged or their signage is faded. Next, using a new shape classification algorithm based on Haar-like features, the candidates are further refined and categorized based on their shape. Object detection and classification problem is traditionally solved by either the selective extraction of windows of interest, or exhaustive sliding window based classification (Balali et al. 2013). In the first approach small number of interest regions are selected in the images through fast and inexpensive methods. These interest regions are then subjected to a more sophisticated classification. Such approach risks overlooking some traffic signs. Second approach considers all candidate windows in the image. Given the large number of candidates, classification easily becomes intractable.

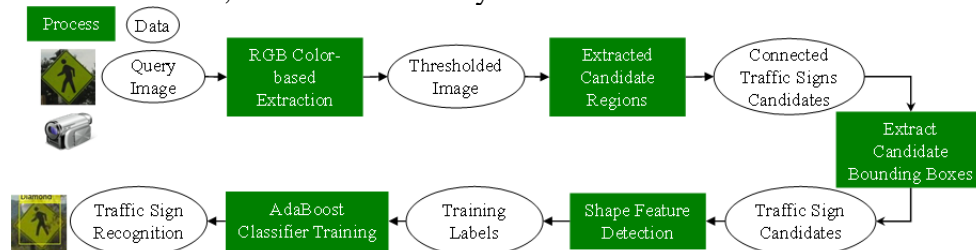


Figure 2. System overview for detection and classification process.

Color-based Candidate Detection. Traffic sign colors are easily distinguishable from the background environment. Thus, the first step could be a computationally efficient optimization to select bounding boxes of possible traffic signs from 2D images with a simple color thresholding method; i.e., every pixel with values above a

threshold is marked with the appropriate label. The thresholded image is obtained from an RGB color video frame by applying threshold $T = (t, \alpha, \beta, \lambda)$ at the pixel level similar to (Timofte et al. 2011):

$$V(T) = \begin{cases} 1 & \alpha.V_R + \beta.V_G + \lambda.V_B \geq t \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Since there is no single threshold performing well by itself, it is necessary to combine regions selected by different thresholds ($\rho = \{T_1, T_2, T_3, \dots, T_n\}$) in order to add Operational Researching (OR). A set of pixels passed on by any threshold are initially selected as candidate location of traffic signs. The more thresholds are used, the lower FN can be made. This however can cause higher rates of FP and consequently higher computational time. Given thousands of possible color thresholds, the optimal subset of (ρ) is selected subject to a constraint. Here, we use a straight forward constraint, a trade-off between FPs and FNs:

$$\rho^* = \arg \min_{\rho} (FP(\rho) + w.FN(\rho)) \quad (2)$$

where $FP(\rho)$ stands for the number of false positives and $FN(\rho)$ for the number of false negatives of the selected thresholds (ρ) respectively. The real number w is a relative factor for weighting. This optimization problem is solved as a Boolean Linear Programming problem. Then, using the Maximally Stable Extremal Region (MSER) feature detection technique (Matas et al. 2004), the pixels are grouped in several bounding boxes as traffic sign candidates. Ultimately a set of initial bounding boxes are extracted wherein each bounding box potentially includes a traffic sign. While this algorithm returns very few FN (non-detected traffic sign), it is designed to return a high number of FP (potential candidates for traffic sign).

Shape-based Candidate Classification. Traffic signs have specific characteristic shapes. In this step, the initial candidates are refined by using a shape classification algorithm which filters out the FPs of the previous color-based detector based on Haar-like features and ultimately categories traffic signs into categories of similar shapes (e.g., rectangle, diamond). (Viola and Jones 2001) developed an algorithm capable for detecting objects. The detector is trained using a set of positive and negative samples.

In our proposed method, by first using a sliding window detector, the selected regions from image candidates are convolved with multiple Haar-like filters. Our detector similar to (Viola and Jones 2001) is a cascade of boosted Haar-like classifiers which uses AdaBoost learning method. The pixel intensities of adjacent rectangular regions are summed up and the differences are calculated from the Haar-like features. Six feature types used in this paper are shown in Figure 3. These features are calculated very fast and are independent of different image resolutions and are robust to noise and changes in illumination. They also can be easily scaled to detect objects at other scales than in the training images.

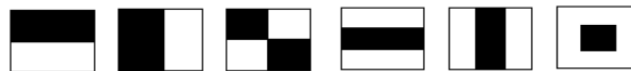


Figure 3. Haar-like features used in our implementation.

Vision Cascade Detector. The vision cascade object detector detects traffic signs in images by sliding a window over the image candidates. The detector then uses a cascade classifier to decide whether the window contains the traffic sign. The size of the window varies to detect traffic signs at different scales, but its aspect ratio remains fixed (1:1). The detector is very sensitive to out-of-plane rotation, because the aspect ratio changes for most traffic signs. Thus, there is a need to train a detector for each orientation.

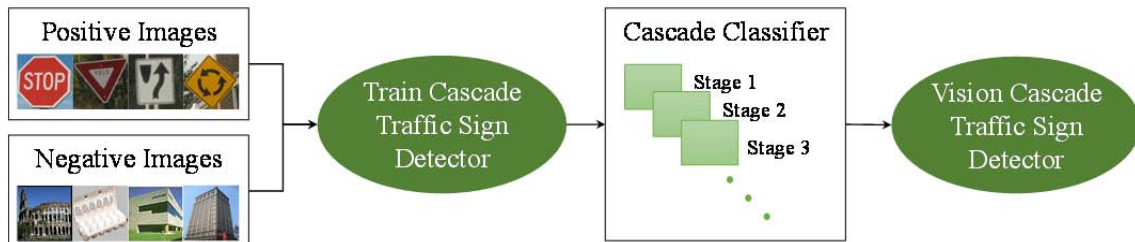


Figure 4. Vision cascade traffic sign and mile marker detector.

Cascade Classifier. Each stage of the classifier labels the region defined by the current location of the sliding window as either positive or negative. If the label is negative, the classification of this region is complete, and the detector slides the window to the next location. If the label is positive, the classifier passes the region to the next stage. The detector reports an object found at the current window location when the final stage classifies the region as positive.

These steps are designed to quickly reject negative samples. The assumption is vast majority of windows do not contain traffic signs. Conversely, True Positives (TP) are rare, and worth taking the time to verify. To work well, each stage in the cascade must have a low FN rate. If a stage incorrectly labels a traffic sign as negative, the classification stops, and there is no way to correct the mistake. However, each stage may have a high FP rate. Even if it incorrectly labels a non-traffic sign as positive, the mistake can be corrected by subsequent stages. The overall FP rate of the cascade classifier is (f^s) , where f is the FP rate per stage in the range [0 1], and s is the number of stages. Similarly, the overall TP rate is (t^s) , where t is the TP rate per stage in the range [0 1]. Thus, adding more stages reduces the overall FP rate, but it also reduces the overall TP rate. Figure 5 shows the process of cascade detector training.

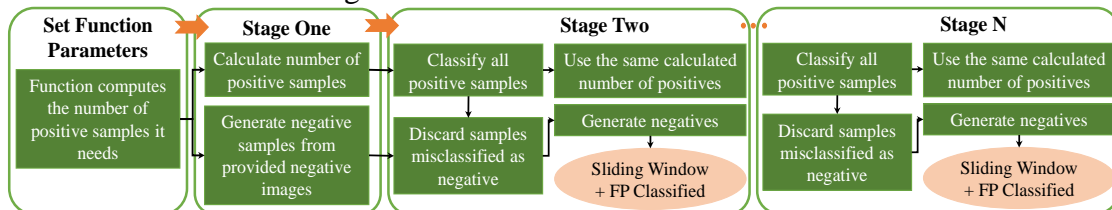


Figure 5. Train cascade traffic sign and mile marker detector.

Sliding Window Comparison. The outlined pipeline, is implemented with different scale factors of sliding window on image candidates: (0.75, 1.00, 1.25) with

6.67% shift – i.e., window overlap. Table 1 shows the parameter values in our cascade detector.

Table 1. Cascade Detector Parameters.





Parameter	Value
False Alarm Rate	0.2
Number of Cascade Stages	5
True Positive Rate	0.995

The false alarm rate is the fraction of negative training samples incorrectly classified as positive samples. Increasing the number of stages may result in a more accurate detector but also increase training time and requires more training images. The true positive rate is the fraction of correctly classified positive training samples.

EXPERIMENTAL RESULTS AND DISCUSSIONS

For evaluating the performance of the traffic signs detection and classification, the experimental dataset was collected on US-460 and secondary roads in a corridor of 8 miles. This dataset is used for both training and testing purposes with split of 70/30. The data is compiled from 30min recorded video using a single camera pointing towards the right side of the road at speed of 40 km/h. The dataset contains different shapes of traffic signs which are annotated manually for training purpose. To increase negative samples needed for training AdaBoost classifier, we also added 16,000 negative samples of typical sign backgrounds (see Table 2). The entire dataset is split into different types of signs based on shape and corresponding number of positive annotations and negative samples.

Table 2. Summary Information of Traffic Signs and Mile Markers.

Type	Color	#Positives	#Negatives	Sign Message	
Diamond	Yellow	2280	27758	Warning	
Rectangle	White, Blue, Green	7580	33058	Regulatory, Direction (including mile markers)	
Hexagonal	Red	235	10017	Stop	
Triangle	Red	157	10095	Yield, Slow down, Prepare to stop	

The validation process of *color-based detection* is shown in Figure 6 which consists of implementing the thresholded frame extraction method and then selecting the potential frames that contains traffic signs.

The results of different sliding window sizes for different type of traffic signs is shown in Table 3. As it shows, the 64x64 pixel image patches have the minimum FN rates and maximum FP rates among all other sliding window sizes. Figure 7 shows some examples of FNs and FPs of sliding window with size of 64x64 pixels.

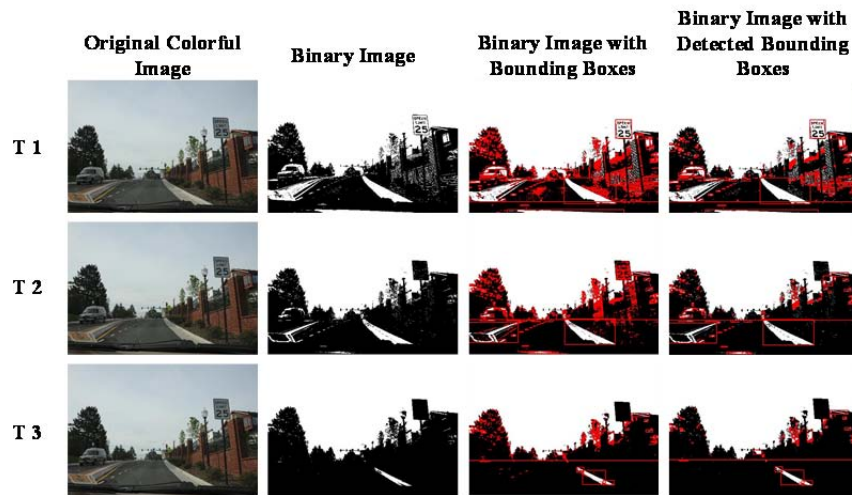


Figure 6. Color-based extraction method for optimum thresholds.

Table 3. Results of FP and FN Rates of Different Size of Sliding Window.

Scale Factor	False Negatives			False Positives		
	0.75	1.00	1.25	0.75	1.00	1.25
Diamond	0.06%	0.06%	3.64%	13.68%	14.19%	36.76%
Rectangle	3.67%	2.17%	6.48%	11.19%	19.23%	21.95%
Hexagonal	0.00%	0.00%	9.23%	23.08%	30.77%	23.08%
Triangle	0.00%	0.00%	4.49%	14.61%	29.21%	38.20%



Figure 7. Examples of FNs and FPs.

Figure 8 shows cases that could be detected for different type of traffic signs and the results of the candidate classification (TP). The shape extraction significantly increases the number of FPs. Precision and recall metric is used to measure the accuracy of classification for sliding window size 64×64 for different types of traffic signs (See Table 4). The average precision and recall among all types of traffic signs is 76.65% and 99.33% respectively. The accuracy of classification is calculated based on $\frac{TP + TN}{TP + TN + FP + FN}$ which is 79.30%.



Figure 8. Examples of true positives.

Table 4. Precision and Recall for Different Types of Traffic Signs.

Type	Precision	Recall	Type	Precision	Recall
Diamond	85.81%	99.93%	Hexagonal	69.23%	100.00%
Rectangle	80.77%	97.38%	Triangle	70.79%	100.00%

CONCLUSION

This paper presented and validated a video-based method for detection and recognition of multi-class of traffic signs which has potential to provide quick and inexpensive access to information about location and condition of highway assets. It can also foster information sharing and exchange among different agencies as well as DOTs. Ongoing efforts focus on detailed validation of each part of the proposed method, and also testing the scalability of the overall method on large datasets from interstate highways in the state of Illinois. These results will be presented soon.

REFERENCES

- (FHWA), F. H. A. (2003). "Manual on uniform traffic control devices (MUTCD)." U.S. Department of Transportation, Washington, D.C.
- Balali, V., and Golparvar-Fard, M. (2014). "Scalable Non-Parametric Parsing for Segmentation and 3D Localization of High-Quantity Low-Cost Highway Assets from Car-Mounted Video Streams." (*CRC*), Atlanta, GA, USA.
- Balali, V., Golparvar-Fard, M., and de la Garza, J. (2013). "Video-Based Highway Asset Recognition and 3D Localization." *Computing in Civil Eng.*, 379-386.
- Brkic, K. (2013). "An overview of traffic sign detection methods." Electrical Engineering and Computing, Unska, Zagreb, Croatia, 10000.
- Cheok, G., Franaszek, M., Katz, I., Lytle, A., Saidi, K., and Scott, N. (2010). "Assessing Technology Gaps for the Federal Highway Administration Digital Highway Measurement Program.", National Institute of Standards and Technology (NIST).
- Fatmehsan, Y. R., Ghahari, A., and Zoroofi, R. A. (2010). "Gabor wavelet for road sign detection and recognition using a hybrid classifier." *MCIT*, 25-28.
- Golparvar-Fard, M., Balali, V., and de la Garza, J. (2012). "Segmentation and Recognition of Highway Assets Using Image-Based 3D Point Clouds and Semantic Texton Forests." *Computing in Civil Eng.*, 10.1061/(ASCE)CP.1943-5487.0000283.
- Hu, Z., and Tsai, Y. (2011). "Generalized Image Recognition Algorithm for Sign Inventory." *Journal of Computing in Civil Engineering*, 25(2), 149-158.
- Kim, J., Jung, K., and Hyun, C. (2005). "A Study on an Efficient Sign Recognition Algorithm for a Ubiquitous Traffic System." *DSP Computational Science and Its Applications*, Springer Berlin / Heidelberg, 239-246.
- Liu, Y. X., Ikenaga, T., and Goto, S. (2006). "Geometrical, Physical and Text/Symbol Analysis Based Approach of Sign Detection." *Intelligent Vehicles Symp*, 238-243.
- Matas, J., Chum, O., Urban, M., and Pajdla, T. (2004). "Robust wide-baseline stereo from maximally stable extremal regions." *Image and Vision Com.*, 22(10), 761-767.
- Timofte, R., Zimmermann, K., and Van Gool, L. (2011). "Multi-view traffic sign detection, recognition, and 3D localisation." *Machine Vision and Applications*, 1-15.
- Tsai, Y., Kim, P., and Wang, Z. (2009). "Generalized Traffic Sign Detection Model for Developing a Sign Inventory." *Computing in Civil Engineering*, 23(5), 266-276.
- Viola, P., and Jones, M. (2001). "Robust real-time object detection." *J. Computer Vision*, 4.
- Wu, J., and Tsai, Y. (2006). "Enhanced Roadway Inventory Using a 2-D Sign Video Recognition Algorithm." *Computer-Aided Civil and Infrastr Eng*, 21(5), 369-382.

Zhu, S.-d., Zhany, Y., and Lu, X.-f. (2005). "Detection for triangle traffic sign based on neural network." *Proc., Int. Conference on Vehicular Electronics and Safety*, 25-28.