Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Traffic sign detection for U.S. roads

Møgelmose, Andreas; Liu, Dongran; Trivedi, Mohan M.

Traffic Sign Detection for U.S. Roads:
Remaining Challenges and a case for Tracking

Andreas Møgelmose[1,2], Dongran Liu[2], and Mohan M. Trivedi[2]

Abstract— Traffic sign detection is crucial in intelligent vehicles, no matter if one's objective is to develop Advanced Driver Assistance Systems or autonomous cars. Recent advances in traffic sign detection, especially the great effort put into the competition German Traffic Sign Detection Benchmark, have given rise to very reliable detection systems when tested on European signs. The U.S., however, has a rather different approach to traffic sign design. This paper evaluates whether a current state-of-the-art traffic sign detector is useful for American signs. We find that for colorful, distinctively shaped signs, Integral Channel Features work well, but it fails on the large superclass of speed limit signs and similar designs.

We also introduce an extension to the largest public dataset of American signs, the LISA Traffic Sign Dataset, and present an evaluation of tracking in the context of sign detection. We show that tracking essentially suppresses all false positives in our test set, and argue that in order to be useful for higher level analysis, any traffic sign detection system should contain tracking.

I. INTRODUCTION

Advanced Driver Assistance Systems (ADAS) and autonomous cars are currently very popular research topics, with many applications, as car manufacturers continue to put efforts into making cars safer and easier to drive. In both scenarios it is crucial for the onboard systems to have a solid view and understanding of the world around the car.

Part of this understanding comes from being able to read traffic signs. Whether the aim of the system is to assist the driver, or remove him completely from the equation, traffic signs provide valuable information about the driving scenario at any given time. Traffic Sign Recognition (TSR) has seen considerable work over the past decade. The task is generally split into two: detection and recognition [1]. Some full-flow systems of course combine the two tasks, but generally as two separate steps in a pipeline. Other systems focus on one or the other.

This work is focused on detection, and particularly detection of American traffic signs. A recent competition called the German Traffic Sign Detection Benchmark (GTSDB) [2] had several excellent entries, and it was even suggested by one of the entries, Mathias et. al. [3], that the problem was largely solved. Indeed Mathias et. al., along with other top contenders in the competition, showed very impressive detection results on the GTSDB. However, as the name suggests, in includes only German traffic signs, which are substantially different from those in the United States.

[1]Visual Analysis of People Lab, Aalborg University, Denmark `am@create.aau.dk`
[2]Laboratory for Intelligent and Safe Automobiles, UC San Diego, United States `[dol060, mtrivedi]@ucsd.edu`

The purpose of this paper is threefold:
1) Test the method successfully employed in [3] on American signs in order to ascertain whether the traffic sign detection problem is really solved.
2) Add a second layer of intelligence to traffic sign detection in the form of tracking.
3) Introduce a recently captured extension of the LISA Traffic Sign Dataset which adds more high-resolution color imagery to what is already the world's largest and only public dataset of American traffic signs.

The remainder of this paper is structured as follows: In section II, related work is briefly discussed. Section III explains the methods used for both detection and tracking. In section IV the results are shown and discussed for several traffic sign classes, and finally the paper is rounded off with conclusions in section V.

II. RELATED WORK

Traffic sign detection has been heavily researched in the past decade. A comprehensive overview is given in the survey [1]. There are two main approaches: Model-based and learning-based. The earliest approaches [4]–[7] were model-based and relied on static knowledge of traffic sign appearance, such as shape or color. This works well under good conditions, reliable lighting, etc. However, the learning-based methods have taken over lately, since they cope better with real-world variations in the input data. They use a wealth of features, such as Haar-wavelets [8], HOG [9], or Integral Channel Features [3].

Traffic sign classification has a history of being improved through competition, and in 2011, the problem was generally considered solved with the publication of the results of the German Traffic Sign Recognition Benchmark (GTSRB, not to be confused with GTSDB) [10]. Recently, the same authors endeavored to repeat the success for detection with the GTSDB [2]. They provided a dataset of 900 images captured in Germany containing 1206 traffic signs in four categories: prohibitory signs, danger signs, mandatory signs, and other signs. A baseline detection performance was computed using standard methods, and the community was asked to contribute their best detection results.

18 teams contributed their results, generally with very good results. The best competitors were *Team Litsi* [11], *Team VISICS* [3], and *Team wgy@HIT501* [12]. *Team Litsi* used color classification and shape matching to finds regions of interest and detect signs using HOG and color histograms classified with support vector machines. *Team VISICS* use the Integral Channel Features (*ChnFtrs*) proposed by Dollár et.

al. [13] for pedestrian detection. *ChnFtrs* use a combination of edge features and LUV color features in a boosted cascade. *Team wgy@HIT501* uses HOG features, first with LDA to find sign candidates and then HOG with IK-SVM for further refinement. It is interesting to see that the most successful detectors all use detection schemes adapted from pedestrian detection. Note that all of the above work has been done using European traffic signs. Only a few works explicitly take on American signs: In [14] a shape based detector and classifier is used for American speed limit signs. Detection results are not shown on their own, but the final recognition results are around 88% recognition rate, with no mention of the number of false detections. The system is tested on a relatively small dataset. [15] evaluates whether synthetic training data is useful for sign detection, but without great success. Finally, [16], [17] propose an adaptive Bayesian Classifier Cascade, again for speed limit sign detection. It shows promising detection results in the 90%-range, albeit with several false positives per image, much higher than what would be acceptable in a production system. Older contributions include [8], [17], [18].

Little work has been done to extend the pipeline further than raw detection and classification. A few groups work on traffic sign inventory and surveying, including [19], [20]. For any system which aspires to use TSR in a practical system - be it ADAS, inventory systems or autonomous driving - tracking is beneficial, even necessary [1], but building tracking on top of detections has not been thoroughly researched, except very recently, where Boumediene et. al. proposed tracking traffic signs using the Transferable Belief Model (TBM) [21].

This work takes *ChnFtrs*, the method employed by *Team VISICS*, and applies it to American traffic signs. On top of the detection we build an association and tracking layer using Kalman filters.

III. METHODS

The overall structure of the system described in the paper is simple: Detection using *ChnFtrs* and tracking using Kalman filters. In the following, the implementation and use of each method it described in further detail. Everything explained here has been implemented in Matlab.

A. Detection

1) Features: The detection scheme used is Integral Channel Features/*ChnFtrs* by Dollár et. al. [13] as also used by Mathias et. al. [3] in their winning entry in the GTSDB. This method works by computing features similar to Haar-like features in the integral image of several "channels" of the input image. Channel refers to different representations of the input image. We use 10 different channels spanning 6 gradient histogram channels, 1 for unoriented gradient magnitude, and 3 for the color channels in the CIE-LUV color space. The oriented gradient channels are computed as:

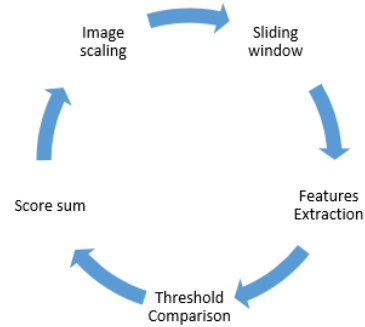$$Q_\theta(x, y) = G(x, y) \cdot \mathbf{1}[\Theta(x, y) = \theta] \qquad (1)$$

Fig. 1. The flow of the detection system.

where $G(x, y)$ is the gradient magnitude at pixel x, y and $\Theta(x, y)$ is the quantized gradient angle. $Q_\theta(x, y)$ is then run for 6 equally spaced angles, θ.

The features we use are the same for all channels, namely first-order Haar-like features. Traditional Haar-like features are expressed as the difference of the sum of two rectangular image regions. First-order Haar-like features are even simpler and expressed as the sum of a rectangular image region in a given channel. The *ChnFtrs* framework allow for more complex features, but according to [13], the gain from using higher-order features is negligible. The features can be computed fast from an integral image representation of the relevant channels.

The features are evaluated using a modified AdaBoost classifier with depth-2 decision trees as weak learners.

2) Training: The system is trained with at set of positive training images, all resized to 15x15 pixels, since that is the smallest size of sign we are interested in detecting. Any less, and the features become unreliable. Similarly, a set of negative samples is also input to the system. For a 15x15 px image patch, we generate 3025 rectangular windows for the first-order features. Since there are 10 channels for each training patch, the total number of computed features is 30250 per patch. As opposed to [3], we do not use different aspect ratios in our training, as we observe that to very rarely be a source of missed detections. After extraction, the features are sent to AdaBoost to train the classifier. We use a 3 stage cascade with different number of weak learners per stage. The first stage uses only 10 weak learners to quickly discard a lot of non-traffic sign windows. The second stage has 100 weak learners and the last stage uses up to 2000 weak learners to refine the detection. However, training of the last stage usually stops early at 300-600 weak learners due to convergence.

3) Detection: We slide a 15x15 px window across the integral image of each channel in the test image. Like in the original *ChnFtrs*, the window is moved by 4 pixels for each evaluation and to ensure scale invariability, the input image is scaled by a step of $2^{1/10}$ (1.07).

An overview of the detection flow is shown in fig. 2

	DR	FPPF
Our implementation	93%	0.07
Team VISICS [3]	100%	0

B. Tracking

Tracking is built on top of the detection. It has two purposes: It can be used to minimize false detections by suppressing those which are not part of a track, and more importantly, they can keep track of which signs have been seen before. In real-world use of a sign detector, we want to ensure that every sign is seen, but also that no sign is seen more than once. Since each specific sign relates to a specific traffic situation, the system needs to know whether that situation has been handled or not, and not blindly repeat itself when a new instance of the same physical sign is detected.

The tracking is handled using the Hungarian algorithm for assignment and Kalman filters for tracking. Every time a detection happens, it is either assigned to an existing track, or a new track is created for it. Any existing track which does not get a sign assigned to them are marked as invisible. If a track is visible for more than 3 frames, a new detection will be announced. Any track which does not last for 3 frames is considered a false detection. Conversely, if a track is invisible for more than 2 frames, or more than 40% of its contained frames are marked invisible, it is deleted.

IV. EVALUATIONS OF DETECTION AND TRACKING

We have two separate evaluations: Firstly we test the detection scheme on American signs, secondly we show the advantages of employing tracking for traffic sign detection.

A. Detection

No matter the test set or sign type, throughout this section we report two numbers per test:

DR Detection rate, computed as TP/P where TP is the number of correct detections in a frame (true positive) and P is the actual number of signs in the frame (positives). This is equivalent to recall.

FPPF False positives per frame, computed as FP/f where FP is the number of false positives across all frames, and f is the number of frames analyzed.

While this paper is about detecting American traffic signs, we first test our system on the GTSDB superset of prohibitory signs as a sanity check to verify that our implementation is not faulty. The result can be seen in table 3. While our performance is significantly lower than that of [3] - perfect detection is hard to beat - it is still respectably above 90%. The difference is probably due to varying parameter settings between the two implementations.

With this baseline check out of the way, we turn our attention to American signs. Specifically, we test on three superclasses: Stop signs, warning signs (all yellow diamond

Fig. 2. Examples of american traffic signs. (a) Stop sign, (b) warning sign, (c) speed limit sign. Image source: [22]

	LISA Traffic Sign Dataset	LISA Dataset Extension
Classes	49	15
Annotations	7855	1326
Images	6610	1445

shaped classes), and speed limit signs (all speed limit signs, regardless of speed). See fig. 3 for examples.

The detector is trained on the LISA traffic sign dataset [1]. Note that this detection method requires color information and not all images in the LISA dataset are in color, so only the 3156 annotations on color images are used. The detector is tested on a recently captured and annotated extension to the LISA traffic sign dataset, expanding the dataset by more than 1300 annotations. The extension will be integrated with the existing publicly available dataset[1]. Statistics can be seen in table 3.

Detection results for each class can be seen in table 3 and precision-recall curves for each superclass are shown in fig. 4. Stop signs and warning signs both perform well, basically on par with European signs. This is to be expected, as they are quite similar to European signs with distictive shapes and clear colors. For speed limit signs, the story is different. Detection rates a abysimally low. Clearly, the model is not able to capture what distinguished the speed limit signs from everything else. The shape is not particular distinctive – we tried adding a margin to the training images, in the hope that the edges would be detected better then, but to no avail – and the signs have no color that make them stand out.

The problems are also obvious in fig. 4 and 5, showing false detection and misses for each superclass. False detections of stop signs and warning signs are generally objects which bear some resemblance to the signs. Missed detections

---

[1]Available for download at `http://cvrr.ucsd.edu/lisa/lisa-traffic-sign-dataset.html`

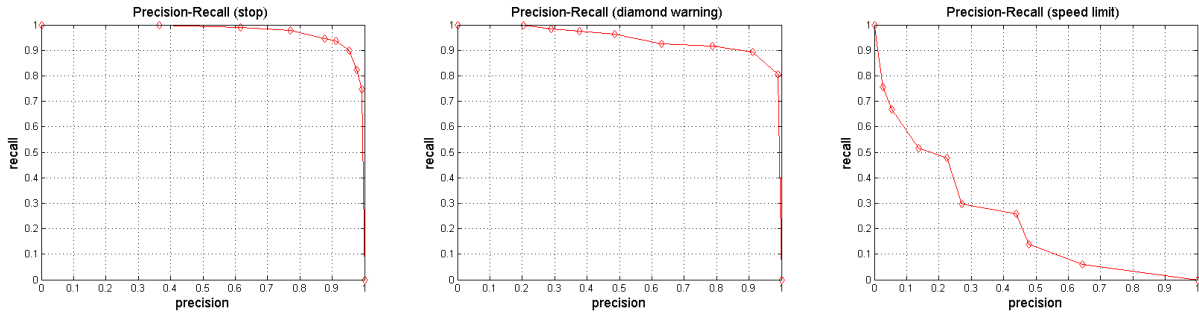|	DR	FPPF	AUC in fig. 3
Stop signs	93.6%	0.09	0.975
Warning signs	89.2%	0.09	0.944
Speed limit signs	47.7%	1.64	0.235

Fig. 3. Precision-recall curves for each superclass. (a) Stop signs, AUC: 0.975, (b) warning signs, AUC: 0.944, (c) speed limit signs, AUC: 0.235.
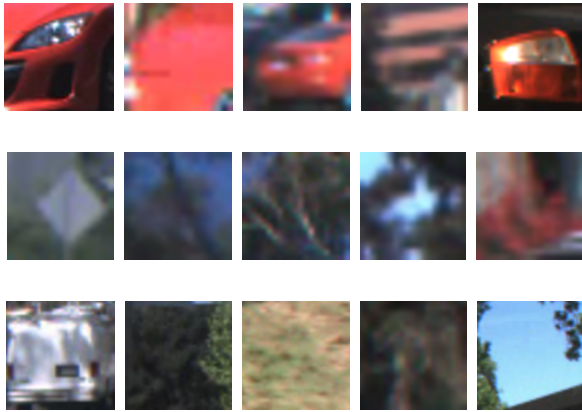

Fig. 4. Examples of false detections for each superclass. First row, stop signs, second row, warning signs, and third row, speed limit signs.
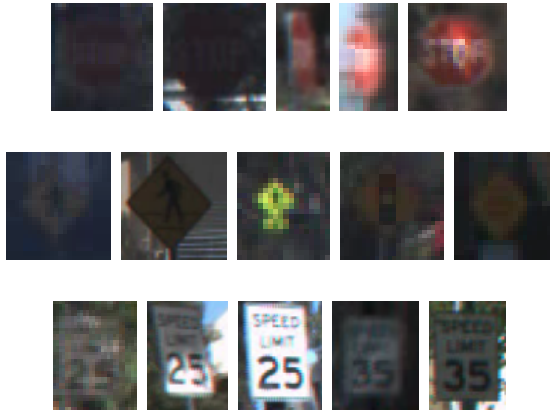

Fig. 5. Examples of missed detections for each superclass. First row, stop signs, second row, warning signs, and third row, speed limit signs.

TABLE IV
TRACKING DETECTION RATES.

	Stop signs	Warning signs
# of physical signs	77	62
Physical signs tracked	76	60
False tracks	2	1
Individual detections	682	597

B. Tracking

The main purpose of our tracking implementation is to make sure the systems know which detections belong to the same physical sign, so each sign is only handled once. A fortunate side effect is that any spurious false detections are also removed, since they do not belong to any track. We have tested it on video sequences of driving taken from the LISA Traffic Sign Dataset Extension. The sequences cover a total of 139 physical signs. The system has been implemented so new tracks show up bottom right of the frame as soon as they are established (after 3 consecutive detections). When they move to a continued tracks in the following frame, they are shown bottom right with a green border. Sample output can be seen in fig. 5.

Actual statistics for the tracking performance can be seen in table 4. The system scores well. Nearly all physical signs are seen and tracked, and only 3 false tracks appear. In other words, almost all false detections are successfully suppressed. We did not run this test on speed limit signs, as the raw detection rate is simply too low for those. This clearly shows the tremendous effect tracking can have on any traffic sign detection systems.

V. CONCLUSION

This work evaluated a state-of-the-art traffic sign detector on American traffic signs. We implemented *ChnFtrs* and trained it on three superclasses of traffic signs: stop, warning, and speed limit. The detector is known to perform extremely well on European traffic signs, so in order to ascertain whether the challenge of detecting traffic signs is truly solved. Our research showed that while the method performs well for colorful, distinctive traffic signs, such as the superclasses stop and warning, it is not able to detect speed limit signs with any acceptable certainty. As part of

are generally signs of very poor quality, either due to size or lighting. That is not as clear cut for speed limit signs. Many of the false detections could be anything, and the missed signs are perfectly fine speed limit signs.

This leads us to conclude, that while sign detection might be solved for European signs and other "easy" signs, there is still plenty of work to be done for hard cases, such as American speed limit signs. And then we have not even touched the multitude of other American signs with the exact same shape and color as the speed limit signs.

Fig. 6. Examples of tracked signs. A sign icon to the bottom left means it has just been established as a track. An icon bottom right with a green border shows a continued track. The two shown images are consecutive frames.

the testing, we introduced an extension of the LISA Traffic Sign Dataset which adds high-resolution color images and sequences to the existing dataset.

Furthermore, we implemented a Kalman filter based tracking system for traffic signs and showed that in real world use, a good tracking system is essential for any traffic sign detection system - and probably more important than achieving the last 2-3 percentage points towards a perfect detection system.

The most important and difficult challenge is undoubtedly to develop a reliable speed limit sign detector. Apart from speed limit signs, the US has a wide range of other signs following the same template, and no system is able to process those at the moment.

REFERENCES

[1] A. Møgelmose, M. M. Trivedi, and T. B. Moeslund, "Vision based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. Special Issue on MLFTSR, no. 13.4, dec 2012.

[2] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, "Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark," in *Neural Networks (IJCNN), The 2013 International Joint Conference on*. IEEE, 2013, pp. 1–8.

[3] M. Mathias, R. Timofte, R. Benenson, and L. Van Gool, "Traffic sign recognition — How far are we from the solution?" in *Neural Networks (IJCNN), The 2013 International Joint Conference on*. IEEE, 2013, pp. 1–8.

[4] A. Ruta, Y. Li, and X. Liu, "Towards real-time traffic sign recognition by class-specific discriminative features," in *Proc. of the 18th British Machine Vision Conference*, vol. 1, 2007, pp. 399–408.

[5] N. Barnes and G. Loy, "Real-time regular polygonal sign detection," in *Field and Service Robotics*. Springer, 2006, pp. 55–66.

[6] P. Gil Jiménez, S. Bascón, H. Moreno, S. Arroyo, and F. Ferreras, "Traffic sign shape classification and localization based on the normalized FFT of the signature of blobs and 2D homographies," *Signal Processing*, vol. 88, no. 12, pp. 2943–2955, 2008.

[7] R. Timofte, K. Zimmermann, and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3d localisation," in *Applications of Computer Vision (WACV), 2009 Workshop on*. Ieee, 2009, pp. 1–8.

[8] C. Keller, C. Sprunk, C. Bahlmann, J. Giebel, and G. Baratoff, "Real-time recognition of U.S. speed signs," in *Intelligent Vehicles Symposium, IEEE*, june 2008, pp. 518–523.

[9] G. Overett and L. Petersson, "Large scale sign detection using HOG feature variants," in *Intelligent Vehicles Symposium (IV), 2011 IEEE*, june 2011, pp. 326–331.

[10] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German Traffic Sign Recognition Benchmark: A multi-class classification competition," in *Neural Networks (IJCNN), The 2011 International Joint Conference on*. IEEE, 2011, pp. 1453–1460. [Online]. Available: http://benchmark.ini.rub.de/?section=gtsrb

[11] M. Liang, M. Yuan, X. Hu, J. Li, and H. Liu, "Traffic sign detection by ROI extraction and histogram features-based recognition," in *Neural Networks (IJCNN), The 2013 International Joint Conference on*, Aug 2013, pp. 1–8.

[12] S. Wang and J. Zhang, "A New Edge Feature For Head-Shoulder Detection," in *ICIP*. IEEE, 2013.

[13] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral Channel Features," in *BMVC*, vol. 2, no. 3, 2009, p. 5.

[14] J. Abukhait, I. Zyout, and A. M. Mansour, "Speed Sign Recognition using Shape-based Features," *International Journal of Computer Applications*, vol. 84, no. 15, pp. 31–37, 2013.

[15] A. Møgelmose, M. M. Trivedi, and T. B. Moeslund, "Learning to detect traffic signs: Comparative evaluation of synthetic and real-world datasets," in *Pattern Recognition (ICPR), International Conference on*. IEEE, 2012, pp. 3452–3455.

[16] A. Staudenmaier, U. Klauck, U. Kreßel, F. Lindner, and C. Wöhler, "Confidence Measurements for Adaptive Bayes Decision Classifier Cascades and Their Application to US Speed Limit Detection," in *Pattern Recognition*, ser. Lecture Notes in Computer Science, A. Pinz, T. Pock, H. Bischof, and F. Leberl, Eds. Springer Berlin Heidelberg, 2012, vol. 7476, pp. 478–487. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-32717-9_48

[17] *Resource Optimized Cascaded Perceptron Classifiers using Structure Tensor Features for US Speed Limit Detection*, vol. 12, 2011.

[18] F. Moutarde, A. Bargeton, A. Herbin, and L. Chanussot, "Robust on-vehicle real-time visual detection of American and European speed limit signs, with a modular Traffic Signs Recognition system," in *Intelligent Vehicles Symposium*. IEEE, 2007, pp. 1122–1126.

[19] S. Lafuente-Arroyo, S. Salcedo-Sanz, S. Maldonado-Bascón, J. A. Portilla-Figueras, and R. J. López-Sastre, "A decision support system for the automatic management of keep-clear signs based on support vector machines and geographic information systems," *Expert Syst. Appl.*, vol. 37, pp. 767–773, January 2010.

[20] L. Hazelhoff, I. Creusen, and P. de With, "Robust classification system with reliability prediction for semi-automatic traffic-sign inventory systems," in *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*. IEEE, 2013, pp. 125–132.

[21] M. Boumediene, J.-P. Lauffenberger, J. Daniel, and C. Cudel, "Coupled Detection, Association and Tracking for Traffic Sign Recognition," in *Intelligent Vehicles Symposium (IV), 2014 IEEE*, june 2014, pp. 1402–1407.

[22] State of California, Department of Transportation, "California Manual on Uniform Traffic Control Devices for Streets and Highways."