Detection of U.S. Traffic Signs

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

Published in:
I E E E Transactions on Intelligent Transportation Systems

DOI (link to publication from Publisher):
10.1109/TITS.2015.2433019

Publication date:
2015

Document Version
Accepted author manuscript, peer reviewed version

Detection of US Traffic Signs

Andreas Møgelmose, Dongran Liu, and Mohan M. Trivedi, *Life Fellow, IEEE*

***Abstract*—This paper presents a comprehensive research study of the detection of US traffic signs. Until now, the research in Traffic Sign Recognition systems has been centered on European traffic signs, but signs can look very different across different parts of the world, and a system which works well in Europe may indeed not work in the US. We go over the recent advances in traffic sign detection and discuss the differences in signs across the world. Then we present a comprehensive extension to the publicly available LISA-TS traffic sign dataset, almost doubling its size, now with HD-quality footage. The extension is made with testing of tracking sign detection systems in mind, providing videos of traffic sign passes. We apply the Integral Channel Features and Aggregate Channel Features detection methods to US traffic signs and show performance numbers outperforming all previous research on US signs (while also performing similarly to the state of the art on European signs). Integral Channel Features have previously been used successfully for European signs, while Aggregate Channel Features have never been applied to the field of traffic signs. We take a look at the performance differences between the two methods and analyze how they perform on very distinctive signs, as well as white, rectangular signs, which tend to blend into their environment.**

***Index Terms*—Advanced Driver Assistance, active safety, machine vision, traffic signs.**

I. Introduction

TRAFFIC sign detection has become an important topic of attention, not only for researchers in intelligent vehicles and driver assistance areas but also those active in the machine vision area. Traffic Sign Recognition (TSR) generally consists of two layers, detection and classification. With the German Traffic Sign Recognition Benchmark (GTSRB) in 2011, the classification problem was largely solved. To achieve a fully functional TSR system, the detection step needs to work as well. With the introduction of the German Traffic Sign Detection Benchmark (GTSDB) competition, a good amount of work has been done to that effect, even with suggestions of the detection problem being solved [1]. We contend that while good progress has definitely been made, the research community is not quite there yet.

Not all traffic signs look the same, especially the US signs are significantly different in appearance from those in Europe. Systems which do not consider them cannot be expected to perform in the same manner as for what they are designed for - namely almost exclusively European signs. We have taken a fresh look at the specific issues, challenges, features, and evaluation of US traffic signs in a comprehensive manner. To do this in a systematic way, the very first order of business is to draw out differences in how these signs appear. Given these rather stark appearance differences, we undertook a major database collection, annotation, organization, and public distribution effort. Secondly, we explored the overall landscape of appearance based object detection research - including European traffic signs - and carefully selected the two most promising approaches, one (Integral Channel Features) which has offered very good results on European signs and another (Aggregate Channel Features) which was very recently introduced in the literature, but has never been applied to the traffic sign case.

A. Møgelmose is with the Visual Analysis of People Lab, Aalborg University, Denmark, and Laboratory for Intelligent and Safe Automobiles (LISA), UC San Diego, USA. E-mail: am@create.aau.dk.

D. Liu and M. M. Trivedi are with Laboratory for Intelligent and Safe Automobiles (LISA), UC San Diego, USA.

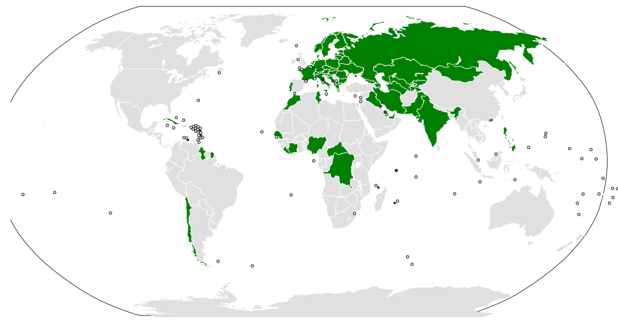


Figure 1. Countries which have ratified the Vienna Convention on Road Signs and Signals. Note that apart from these, Japan, Australia, and to a lesser extent China also follows it, even though they did not ratify it. Data source: [2]

TSR is becoming more and more relevant, as cars obtain better and better Advanced Driver Assistance Systems (ADAS), and driving becomes more and more automated. While mapping-based indexing of traffic signs can replace in-situ recognition to some extent, it will never be able to work in changing road conditions, such as road work, and furthermore the initial sign locations and types must be determined somehow in the first place. Until the infrastructure is updated to include wireless transponders in all traffic signs, TSR will have its place in cars.

No matter the application, detecting and recognizing signs on individual images is not sufficient. If every new detection in a video feed is treated as a new sign, the driver (human or not) will quickly be overwhelmed by notifications. Instead detections must be grouped so all detections pertaining to the same physical sign are treated like the single sign it is. The temporal grouping of detections may also have a positive impact on the classification, since more than one image can be used to determine the sign type. For temporal grouping, tracking comes into play. There has been some research into tracking of traffic signs [3], [4], but it is still in its infancy. One of the issues in traffic sign tracking is that no suitable dataset exists for that purpose, something the dataset extension put forth in this paper addresses, even though we do not tackle that issue in the experiments here.

The primary goal of this paper is to present the most com-

(a) (b) (c)

Figure 2. Examples of Vienna Convention signs. (a) Keep right, superclass mandatory. Sign D15,3. (b) Left turn, superclass danger. Sign A41-2. (c) 60 km/h speed limit, superclass prohibitory. Sign C55.

(a) (b) (c)

Figure 3. US signs corresponding to the Vienna Convention signs in fig. 2. (a) Keep right, superclass traffic movement. Sign R4-7. (b) Left turn, superclass warning. Sign W1-1. (c) 50 mph speed limit, superclass speed limit. Sign R2-1. Image source: [6]

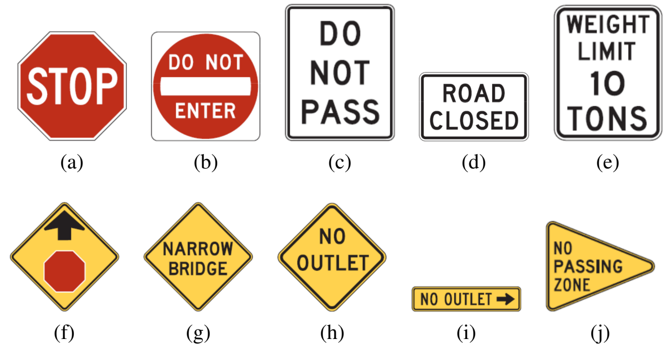
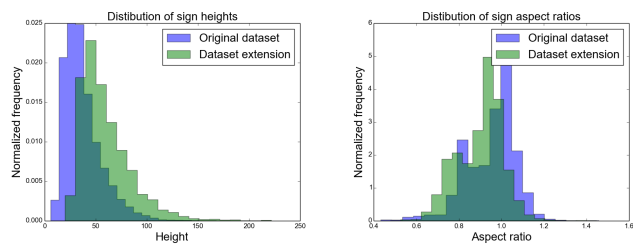


Figure 4. Examples of US signs. (a) Sign R1-1. (b) Sign R5-1. (c) Sign R4-1. (d) Sign R1-2. (e) Sign R12-1. (f) Sign W3-1. (g) Sign W5-2. (h) Sign W14-2. (i) Sign W14-2a. (j) Sign W14-3. Image source: [6]



(a) (b)

Figure 5. Histograms of the (a) heights and (b) aspect ratios of the annotated signs. Statistics for the original dataset and for the extension have been overlaid for easy comparison.

prehensive treatment of the US Traffic Sign Detection studies. Specifically, this paper makes following three contributions:

- We show that while traffic sign detection has indeed come a long way over the last couple of years, international variations in traffic sign designs mean that the problem is by no means solved yet.
- We test two state-of-the-art detection schemes (ICF and ACF) on traffic signs from different countries, with special focus on US signs, which have largely been ignored by the community. We compare their results and achieve state-of-the-art detection results on both European and US signs.
- We introduce a comprehensive extension to the LISA Traffic Sign Dataset [5] which allows for detailed testing of traffic sign tracking.

The paper is structured as follows. In the next section we cover the latest related studies in the field of traffic sign detection. Section III briefly covers what traffic signs are, and especially how traffic sign differ among countries. We also present the extended LISA Traffic Sign Dataset. Section IV describes which detection methods we evaluate. In section V we pit the detection methods against each other.

II. Related Studies

Traffic sign detection has been researched seriously for about a decade. For a general overview, as well as a survey of the research up until 2012, we refer the reader to [5]. Since 2012, the efforts in detection have been stepped up. Following the successes in pedestrian detection, many of those methods have been repurposed for traffic signs. The great catalyst for the recent progress has been the German Traffic Sign Detection Benchmark (GTSDB) [7], which has really pushed the state-of-the-art detection performance to near-perfection on European signs.

Previously, the field was split in model-based and learning-based approaches, but recently the learning-based methods have taken over completely in defining the state-of-the-art. Thus, all the front-runners in the GTSDB competition were learning based. The competition encompassed 18 teams and 3 teams were considered the top performers: *Team VISICS* [1], *Team Litsi* [8], and *Team wgy@HIT501* [9]. *Team VISICS* use the Integral Channel Features (*ChnFtrs* or ICF) proposed by Dollár et al. [10] for pedestrian detection and later improved in [11]. The same method is evaluated in this paper, along with its successor, Aggregate Channel Features. *Team Litsi* first establishes regions of interest with color classification and shape matching and the detect signs using HOG and color histograms with an SVM, features somewhat similar to ICF. Finally, *Team wgy@HIT501* uses HOG features, finding candidates with LDA and performing a more fine-grained detection using HOG with IK-SVM. In essence, all three approaches are rather similar, especially when it comes to features. Another recent paper presenting work on the GTSDB dataset is [12], which shows somewhat worse detection performance than the competitors above, but at a faster speed.

For US traffic signs, the activity has been less enthusiastic. Only a few recent studies take on US traffic signs. In 2012, [13] evaluated whether synthetic training data of US signs could be used successfully to train a rudimentary detector, but performance for the synthetic training data was poor compared to real-world images. Also in 2012, Staudenmaier et al. [14] (building on their previous paper [15]) showed a Bayesian

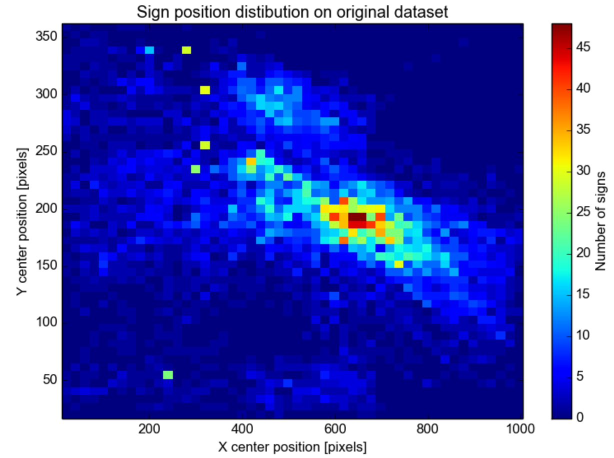
Classifier Cascade with intensity features and tensor features (which describe edges). They detect US speed limit signs at a good rate above 90%, but with several false positives per image, much, much worse than the current European state-of-the-art systems. Abukhait et al. [16] use a shape-based detector to find US speed limit signs - note the model-based, rather than learning-based approach. The detector is part of a full recognition system, and the only reported performance figure is a detection rate of about 88%, but without mention of false positive rates. Stepping back in time to 2008, Keller et al. [17] worked on detection of US speed limit signs using a voting based rectangle detector (see also [18]) followed by an AdaBoost classifier. Moutarde et al. [19], [20] also tackled the case of US speed limit signs using a proprietary rectangle detector. Finally, a precursor to this study was presented in 2014 [4]. To the best of our knowledge, no other US sign based works have been published to date. The existing papers have generally focused on speed limit signs, and achieved significantly worse performance than what we see in the GTSDB.

III. Traffic signs: International conventions and differences

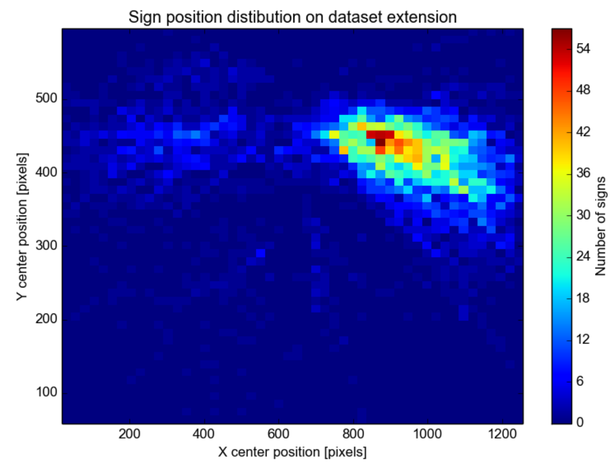
The bulk of the research in TSR systems has laid in European signs, or more specifically signs conforming to the Vienna Convention on Road Signs and Signals [2], [5]. The Vienna Convention has been ratified in 62 countries, as illustrated in fig. 1, so as an initial effort, going after those designs is reasonable. Note that while Australia and Japan have not ratified the convention, they largely design their signs in similar ways to Europe. The same holds true to a lesser extent for China. In other words, the Vienna Convention covers most of Europe and some of Asia, but leaves large parts of the world out, most notably the Americas and south east Asia. Also Africa, though that continent is probably less of a market for ADAS at the moment.

The differences in sign designs matter very much in a detection context. Fig. 2 shows a typical sign from each of the major sign superclasses in Europe: Mandatory, danger, and prohibitory. Each class is very distinctive, not only from the others, but also from most things in the real world. They all have both a rather distinctive shape and a strongly colored border/background. US signs are not in exactly the same classes, but fig. 3 shows the matching US signs. Fig. 4 shows more examples of US signs. From the outset, three things are obvious:

1) US signs bear little to no resemblance to Vienna Convention signs, at the very least requiring re-training of any detector.
2) The strong visual structure in Vienna Convention signs is less present in US signs. The strongest visual clue for US signs is the yellow diamond of warning signs, but even that is not present for all warning signs. The stop sign (which is identical with its Vienna Convention counterpart) is also visually strong. However, most other signs are just white rectangles of varying aspect ratios, which should be challenging to standard detectors which often rely heavily on color cues.



(a)



(b)

Figure 6. Heatmaps showing the positions of annotations in the frame for (a) the original dataset and (b) the extension. The contours of the road can almost be seen in the plots, especially in the extension, which has been captured using an identical camera setup for all frames. The heat maps give a very strong hint towards reasonable Regions of Interest for traffic sign detectors.

3) Many US signs contain only text to convey their message, as opposed to Vienna Convention signs which mostly use icons and pictograms.

Given these large differences to Vienna Convention signs, and the size of the car market in the US, it is surprising that in [5], only two studies were concerned with US traffic signs, and as we describe in the previous section on related studies, not many have come out since the publication of the review.

A. Dataset

In order to properly train and evaluate TSR systems, traffic sign datasets are needed. In [5], the LISA Traffic Sign Dataset was introduced. In this paper we announce a very large extension of the LISA Traffic Sign Dataset, which almost doubles its size with the addition of annotated high-resolution color images. The LISA-TS Extension is split into a training set with every 5th frame annotated and a test set with every

Table I
TRAFFIC SIGN DATASET STATISTICS

	Original LISA Traffic Sign Dataset	**LISA-TS Extension**			**GTSD [7]**	**BTSD [1]**
		Training set	Testing set	Combined		
Number of classes:	47	31	17	35	4	4
Number of annotations:	7855	3672	2422	6094	1206	4627
Number of images:	6610	3307	2331	5638	900	9006
Sign sizes, longest edge:	6 – 168 px	23 – 222 px	25 – 222 px	23 – 222 px	16 – 128 px	16 – 913 px
Image sizes:	640x480 to 1024x522 px	1280x960 px	1280x960 px	1280x960 px	1360x800 px	1628x1236 px
Includes videos:	Yes, but not public	Yes	Yes	Yes	No	Image sequences
Video annotation density:	Every 5 frames	Every 5 frames	All frames	Mixed	N/A	N/A
Notes	Some images in grayscale				The provided classes are actually superclasses.	Classes are superclasses. Multi-view: 8 cameras on one car.

frame annotated, so it is also suitable for testing traffic sign tracking systems. Tracking is outside of the scope of this paper, but see [4] for a simple tracking experiment and [3] for a more advanced solution.

The extension has been collected in and around San Diego, California on urban streets during the spring of 2014. All images were captured with a Point Grey FL3 color camera at the resolution of 1280x960 at approximately 15 fps. Weather conditions are generally dry and sunny or cloudy.

Table I shows statistics about both LISA-TS and LISA-TS Extension, and compares them to the two other relevant detection datasets, the German Traffic Sign Detection set (GTSD) [7] and the Belgium Traffic Sign Detection set (BTSD) [1]. When LISA-TS and LISA-TS Extension are combined, the dataset size exceeds even BTSD.

Fig. 5 shows the size and aspect ratio distributions for the original LISA-TS set and the LISA-TS Extension. Height-wise we see a similar distribution on both datasets, but shifted towards larger sizes for the extension. This fits well with the extension being higher resolution (sizes are measured in pixels) and the similar distributions show that both datasets cover signs at approximately the same distances. With regards to aspect ratio, the distributions are also similar. Many signs have an aspect ratio of 1.0 - this covers round, square and octagonal signs, and the remaining are around 0.8, which fits well with speed limit signs and other rectangular signs. Ratios outside of this is explained by non-orthogonal viewing angles, which can significantly distort the sign in the image plane.

Position heatmaps are shown in fig. 6, one for the original set and one for the extension. In both cases most signs are clearly positioned along the right shoulder of the road, as expected. The original set is somewhat less clearly defined than the extension, undoubtedly because the original set was captured from different vehicles with slightly differing camera positions, whereas the extension has been captured with a single vehicle with a fixed camera position.

Qualitatively, there are some differences between LISA-TS and LISA-TS Extension. LISA-TS predominantly consists of low resolution images, some in grayscale, but was captured over a larger area in Southern California. LISA-TS Extension has consistent high-res color images, all captured from the

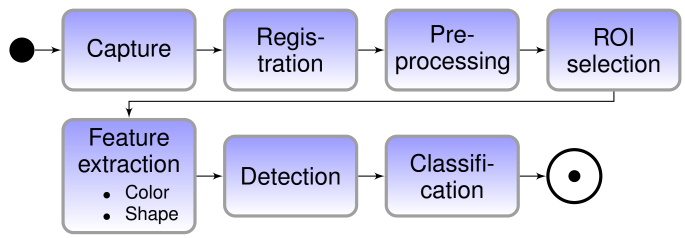


Figure 7. Flowchart of a typical TSR system. A picture is captured and if the systems supports detailed real-world positioning of signs, a registration of this image happens. Afterwards, some pre-processing takes place, usually color normalization followed by an ROI-selection if necessary. This can speed up detection by only looking in relevant parts of the image. Usually the ROI is hardcoded in advance, but saliency measures may also be used. Then relevant features are extracted, most often from a sliding window, detection happens, and finally the detected signs are classified. Some systems contain only some of these blocks, and some systems have this pipeline running in parallel for different sign superclasses.

same rig, ensuring similar images across the set. It also has a dedicated test set. The Extension was captured in and around San Diego. Both datasets suffer from the magnificent sunny Californian weather, so researchers interested in evaluating their algorithms in adverse weather conditions should look elsewhere.

LISA-TS is used as the benchmarking dataset for the sign-detection part of the VIVA 2015 workshop[1] at Intelligent Vehicles Symposium, 2015. For a further overview of other datasets, see [5]. In this paper we use both the GTSD and the expanded LISA-TS dataset.

IV. DETECTION METHODS

We evaluate two state-of-the-art detection methods on traffic sign detection: Integral Channel Features and Aggregate Channel Features. Both are adapted from pedestrian detection and the first has previously been used for traffic sign detection with great success [1]. Our implementation takes its starting point in the Matlab code of Piotr Dollár's Computer Vision Matlab Toolbox [21], with the settings based on his as well. We also discuss image pre-processing, as we find that color normalization is absolutely crucial for good detection

[1]http://cvrr.ucsd.edu/vivachallenge/

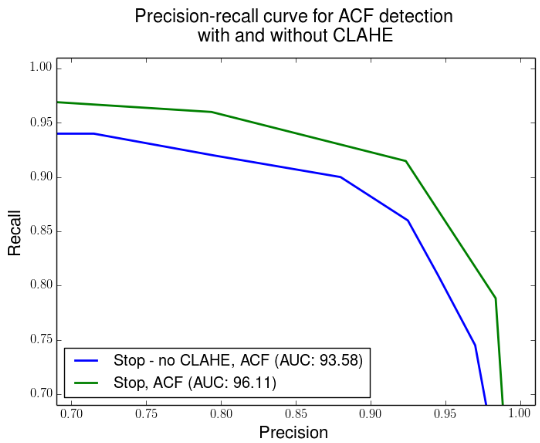


Figure 8. Precision-recall curves for detection of stop signs with and without color pre-processing. The advantage of color pre-processing is evident across the curve.

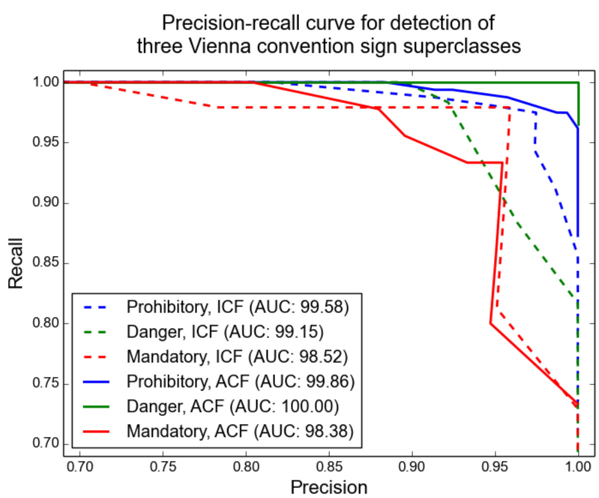


Figure 9. Precision-recall curves for detection of Vienna convention mandatory signs, prohibitory signs, and danger signs. Both ICF and ACF are shown for each superclass. Note that the axes are zoomed to provide a more detailed picture.

performance. Figure 7 shows the flow of a TSR system. Some systems may incorporate just a subset of the blocks shown, but we have included all to give the reader an understanding of the full process. In this paper we focus on capturing (via the dataset capture), pre-processing, feature extraction, and detection. Tracking is not part of the experiments presented here.

A. Image pre-processing

We use contrast-limited adaptive histogram equalization (CLAHE) [22] to normalize colors in the input images. It is a type of histogram equalization which works on tiles in the image, in order to reduce the excessive contrast and noise that may arise from ordinary histogram equalization. Ordinary histogram equalization is done by mapping pixels to a different value based on the cumulative distribution function (CDF) of pixel values in the image. Adaptive Histogram Equalization (AHE), which is slightly simpler than CLAHE, works by performing this transformation for each pixel only by considering the CDF of pixels nearby - a tile. This means that contrast is locally enhanced. A potential problem arises with AHE, though. If a tile is uniform, its CDF has a strong peak, which amplifies pixel noise. CLAHE attempts to combat this by clipping the peak of the CDF so no pixels are too heavily amplified. To speed up computation, the equalization is performed in non-overlapping tiles, which are then blended using bilinear interpolation. Fig. 8 shows the precision-recall curve for stop sign detection with and without CLAHE, clearly demonstrating the importance of this step.

B. Integral Channel Features

Integral Channel Features (ICF, sometimes also abbreviated as ChnFtrs) was first presented by Piotr Dollár for pedestrian detection [10], and recently repurposed by Mathias et al. [1] to achieve near-perfect detection on GTSD. In this paper we put this algorithm to the test on US traffic signs.

The method has two key ingredients: Features computed over 10 different “channels” of the image, and detection using a boosted decision forest. First, the input image is split into 10 channels. The channels used are the three LUV color channels, one for unoriented gradient magnitude, and six for gradients in varying directions. For each of these channels, first-order Haar-like features are computed. This is simply the difference of the sum of two different rectangular image regions. While higher-order features are a possibility, the gain from those is very low. Feature computation is sped up by using the integral image of each channel, C, defined as

$$CC_j(x,y) = \sum_{x' \leq x, y' \leq y} C_j(x',y'), j = 1, \ldots, D \qquad (1)$$

where D is the number of channels. In [1], detectors for each sign are trained for various skewed aspect ratios, to account for non-orthogonal viewing angles. We found only negligible performance gains from this and thus do not have that as part of our training.

After the features are computed, an AdaBoost classifier is learned with depth-2 decision trees as weak learners. This classifier is then run on a sliding window on the input image.

C. Aggregate Channel Features

In 2014, Dollár et al. published an enhanced version of ICF, called Aggregate Channel Features (ACF) [23]. Ostensibly, ACF was introduced as a faster alternative to ICF, but in some cases it shows better detection performance as well. The basic principle about computing features across channels is the same as ICF, and indeed the same channels are used. The Haar-like features are replaced with an even simpler scheme: summing up blocks of pixels at various scales. This is obviously faster than computing the already simple Haar features, but as we shall see provides similar and sometimes even better detection. The boosted decision forest is preserved as the classifier of choice.

Figure 10. Examples of Vienna convention signs which ICF misses. By row: Danger, mandatory and prohibitory.

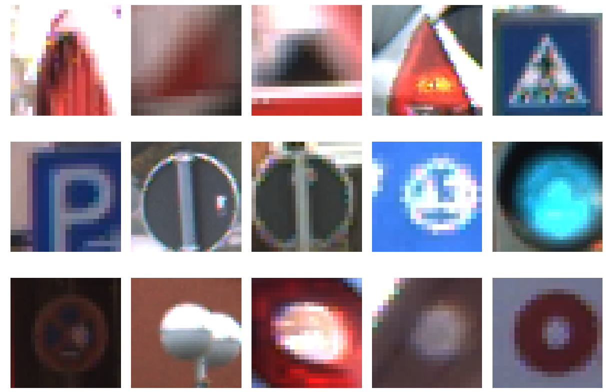


Figure 11. Examples of false positives on Vienna convention signs for ICF. By row: Danger, mandatory and prohibitory.

Table II
LISA-TS TRAINING AND TESTING STATISTICS

	Number of images	
Superclass	**Training**	**Test**
Diamond	1229	406
Stop	1182	1152
NoTurn	185	83
SpeedLimit	750	680

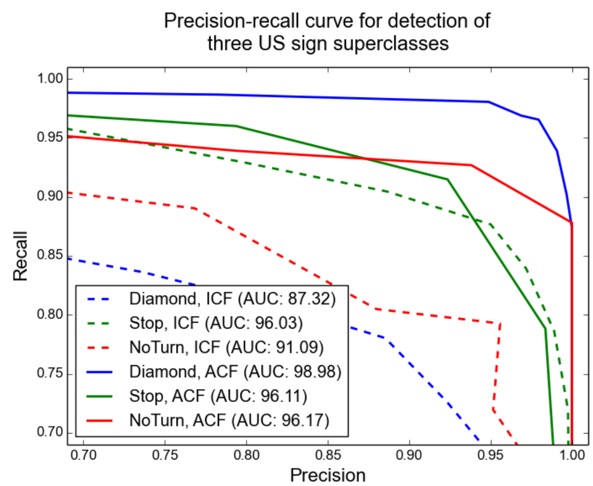


Figure 12. Precision-recall curves for detection of US stop signs, warning signs, and no-turn signs. Both ICF and ACF are shown for each superclass. Note that the axes are zoomed to provide a more detailed picture.

V. EVALUATIONS

We evaluate the two detectors on both the GTSD (to verify that we can replicate the near-perfect results of [1]) and more importantly on the LISA-TS, to show how the methods work on US signs and highlight the challenges unique to US traffic signs. The PASCAL measure [24] has been used to determine detection rates, as is standard. A detection is considered true if the overlap of the detection bounding box and the annotation is more than 50%:

$$a_o \equiv \frac{area(BB_{dt} \cap BB_{gt})}{area(BB_{dt} \cup BB_{gt})} > 0.5 \tag{2}$$

where BB_{dt} is the bounding box of the detection and BB_{gt} the bounding box of the ground truth.

A. European signs

The GTSB is divided into a separate training and test set and spans 4 superclasses: *prohibitory* signs (circular with a red border), *mandatory* signs (circular and blue), *danger* signs (triangular with a red border), and *other* signs, comprising all signs which do not fit in any of the three other categories. *Other* is a very diverse category spanning many shapes and colors, and since it was not considered in the GTSD benchmark, we also ignore it. Results for each of the three superclasses are shown in fig. 9. ACF detects danger signs perfectly - no misses and no false positives. The detection is close to perfect for the remaining two classes as well. Overall, the performance is comparable to that of Mathias et al. in [1]. We fare a little worse on prohibitory signs at an Area Under Curve (AUC) of 99.58/99.86 for ICF/ACF vs. their 100 and a little better on mandatory signs with 98.52/98.38 vs. their 96.98. ICF generally performs very slightly lower than ACF, due to just a few troublesome images, see figures 10 and 11. It is not impossible that both could be further tweaked to obtain perfect scores.

B. US signs

We tested the algorithms on the extended LISA-TS dataset, split up into a training set and a test set. In total, the training set comprise 3346 pictures and the test set contains 2321 pictures. Details for each superclass can be seen in table II.

Detection results for US signs are shown in fig. 12. The precision-recall is generally worse than for European signs, though the diamond superclass just barely surpasses the European mandatory superclass. This shows that even the best performing methods for European signs do not necessarily generalize to other traffic sign design schemes. It is also clear that ACF performs significantly better than ICF for US signs, whereas the difference for European signs was much less pronounced. Diving into the numbers, diamond signs

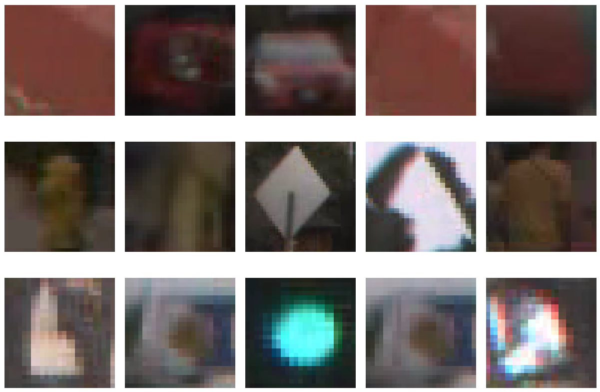


Figure 13. Examples of false positive on US signs. Each row is one superclass: Stop, diamond, and no turn.

Figure 14. Examples of missed US signs. Each row is one superclass: Stop, diamond, and no turn.

are still detected with what can be considered very good performance with an AUC of 98.98 for ACF. The two other superclasses also have an AUC above 95, but there is room for improvement. Examples of false detections and misses can be seen in fig. 13 and 14.

C. US speed limit signs

As shown above, ICF and especially ACF performs very well on European signs and "easy" US superclasses. Common for all these signs is that they have very strong color and shape cues. But a large amount of US signs are not that easy to distinguish (see section III). Many are simply white rectangles. A good representative of this design is the speed limit superclass, shown in fig. 3c. Thus, we dedicate this section to that specific superclass.

We have run both our ICF and ACF detector on this superclass, and indeed the performance is worse. Fig. 15 shows precision-recall curves for both detectors. The other US superclasses have been added for reference. The performance is significantly worse than for the more salient superclasses with AUCs under 90 for both detectors. Still, we note that the performance is much better than the competing US sign detection systems mention in Related Studies (section II). Interestingly, ICF performs slightly better than ACF for this particular superclass. Since there is no color in these signs, the LUV channels used by both detection schemes have very limited impact, other than discarding brightly colored objects. It is also unlikely that a better color normalization scheme will have a significant impact, as was the case for colored signs. Examples of false detections and misses can be seen in fig. 16.

We consider detection of this superclass and its colorless rectangular siblings an open problem. Note also that in this study, we have only looked at speed limit signs, and there is a whole host of other very similar designs. It is not clear from this study whether they should all be lumped together in a monolithic detector, or it is better to have multiple dedicated detectors. While there can be significantly semantic difference for humans between a speed limit sign and a do not pass sign, they have a very similar design, pointing towards the

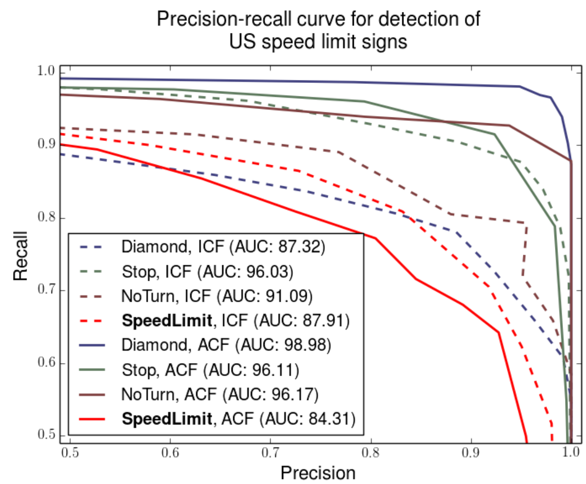
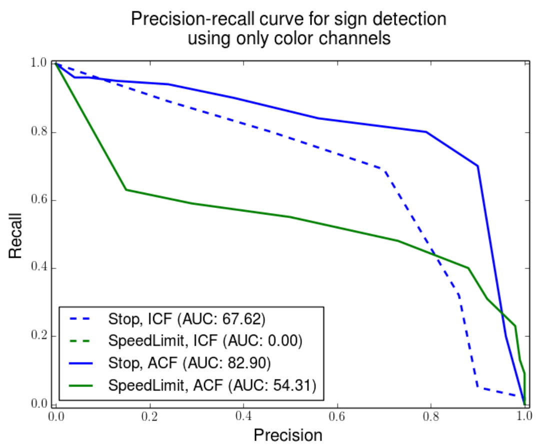


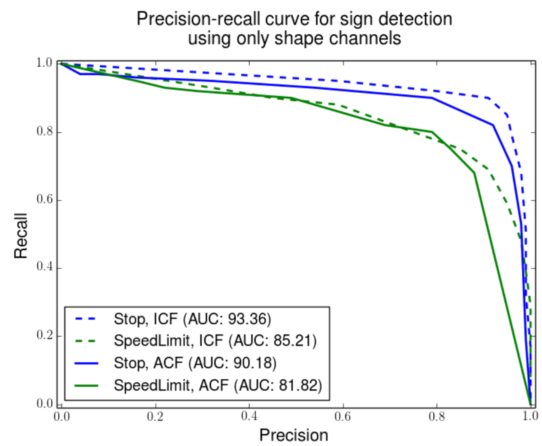
Figure 15. Precision-recall curves for detection of US speed limit signs. For comparison, stop signs, warning signs, and no-turn signs are also shown. Both ICF and ACF are shown for each superclass. The axes are zoomed to provide a more detailed picture, but not as much as for previous figures.

monolithic solution. On the other hand, speed limit signs have very large characters for the numbers, a visual contrast to other text based signs. Furthermore, some signs like that in fig. 3a contain no text at all. It is also worth considering that many of the white rectangular signs have different aspect ratios.

Figure 16. Examples of troublesome US speed limit signs. First row shows missed signs and second row shows false positives.



(a)



(b)

Figure 17. Detection performance using a) only color channels and b) only shape channels for stop and speed limit signs. Shape is a stronger cue.

D. Channel analysis

To better understand the contributions of individual components in detection for different signs, we have run detectors using color channels only and shape channels only (fig. 17). The broad conclusion is that shape is a stronger cue than color. In particular for ICF, where the shape-only detection is close to the performance with all channels for both superclasses - ICF also fails completely on speed limit signs when only color is used. ACF shows a stronger reliance on color cues, though shape is still the most important. This helps explain why ICF performs better than ACF for speed limit signs when all channels are used. Overall, it is interesting to see how differently color and shape contribute in the two methods, given that both methods are building on the same principles. In all cases, though, the combined detector performs better than either of the channel subsets on their own.

Unsurprisingly, color detection works much better on stop signs which have a strong red color, than on speed limit signs, which are just white so the color cue can only be used to discard strongly colored candidate windows.

VI. Concluding remarks

Traffic sign detection has come far on European signs, but not much attention has been given to US signs. This study remedies this discrepancy, by bringing detection of some US sign types - diamond warning signs, stop signs, and no turn signs - up on par with detection rates for European traffic signs with AUCs of 98.98, 96.11. and 96.17, respectively. We test the established ICF detector on US signs and are the first to bring the newer and (in some cases) better ACF algorithm to the domain of traffic sign detection.

Our analysis shows that while we achieve better detection rates on speed limit signs than any studies before us, there is still work to be done on that particular superclass. It is not a given that the methods tested in this paper are the way to go when detecting signs with limited color- and shape cues. Furthermore we provide a large extension to the existing LISA-TS traffic sign dataset, which is publicly available and the only large-scale dataset of US traffic signs in existence.

The most obvious place to direct future research - at least in pure detection - is to push the boundary in detection of US speed limit signs and all the visually related white rectangular signs. Of course tracking is also very relevant for combining distinct detections of the same sign into a single entity, but as indicated by [4], very complicated tracking schemes are perhaps unnecessary in the TSR domain.

A detection method based off of the same methods used here was presented in [25], and it might be possible to adapt to traffic signs and account for detection of traffic signs which are significantly skewed with respect to the image plane. Another interesting aspect could be to combine traffic sign detection with vehicle detection [26] for a more holistic understanding of the traffic situation - for example when a sign signals a lane merge, the position of other cars is very relevant.

Finally, combining a TSR system with a driver attention estimation system, such as in [27], could result in a driver assistance system which dynamically informs the driver only about relevant signs he or she has not seen. This could significantly reduce the rate of signs missed by the driver, while also not causing information overload - as shown in [28] there are whole classes of signs which drivers are very bad at noticing, so such a system would certainly be valuable.

Appendix A
LISA US Traffic Sign Data Set: Expanded Version

The LISA Traffic Sign Dataset is the first and only publicly available data set containing US traffic signs. It was introduced in the 2012 paper *Vision based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey* by Andreas Møgelmose, Mohan M. Trivedi, and Thomas B. Moeslund [5].

It is intended to be used in the development of Traffic Sign Recognition (TSR) systems. The original dataset was captured and released in 2012. In 2014, an extension was made to include higher resolution color images and a dedicated test set where every frame is annotated, to facilitate evaluation of traffic sign tracking systems.

LISA-TS can be downloaded from http://cvrr.ucsd.edu/LISA/lisa-traffic-sign-dataset.html.

Main highlights of the contents and classes of the LISA-TS data set are presented below.

A. Breakdown by class

1) Original dataset:

Superclasses			
3314	warning	1508	speed limit
73	noTurn		
Detailed classes			
294	addedLane	34	slow
37	curveLeft	11	speedLimit15
50	curveRight	349	speedLimit25
35	dip	140	speedLimit30
23	doNotEnter	538	speedLimit35
9	doNotPass	73	speedLimit40
2	intersection	141	speedLimit45
331	keepRight	48	speedLimit50
210	laneEnds	2	speedLimit55
266	merge	74	speedLimit65
47	noLeftTurn	132	speedLimitUrdbl
26	noRightTurn	1821	stop
1085	pedestrianCrossing	168	stopAhead
11	rampSpeedAdvisory20	5	thruMergeLeft
5	rampSpeedAdvisory35	7	thruMergeRight
3	rampSpeedAdvisory40	19	thruTrafficMergeLeft
29	rampSpeedAdvisory45	60	truckSpeedLimit55
16	rampSpeedAdvisory50	32	turnLeft
3	rampSpeedAdvisoryUrdbl	92	turnRight
77	rightLaneMustTurn	236	yield
53	roundabout	57	yieldAhead
133	school	21	zoneAhead25
105	schoolSpeedLimit25	20	zoneAhead45
925	signalAhead		
In total: 7855 sign annotations			

2) Dataset extension: Training:

Superclasses			
1232	warning	752	speed limit
184	noTurn		
Detailed classes			
2	addedLane	91	school
20	bicyclesMayUseFullLane	428	signalAhead
50	curveLeft	4	speedBumpsAhead
59	curveRight	17	speedLimit15
39	doNotEnter	259	speedLimit25
11	intersection	90	speedLimit30
24	intersectionLaneControl	158	speedLimit35
127	keepRight	92	speedLimit40
47	laneEnds	80	speedLimit45
6	leftAndUTurnControl	53	speedLimit50
18	merge	3	speedLimit60
73	noLeftAndUTurn	1181	stop
8	noParking	86	stopAhead
16	noRightTurn	4	yieldAhead
95	noUTurn	8	yieldToPedestrian
523	pedestrianCrossing		
In total: 3672 sign annotations			

3) Dataset extension: Testing:

Superclasses			
461	warning	679	speed limit
82	noTurn		
Detailed classes			
13	curveRight	40	signalAhead
17	dip	406	speedLimit25
21	doNotEnter	264	speedLimit30
11	keepRight	9	speedLimit45
37	merge	1151	stop
29	noLeftTurn	86	stopAhead
53	noUTurn	43	turnRight
80	pedestrianCrossing	145	warningUrdbl
17	school		
In total: 2422 sign annotations			

B. List of superclasses

warning	addedLane	**speedLimit**	speedLimit15
	curveLeft		speedLimit25
	curveRight		speedLimit30
	dip		speedLimit35
	intersection		speedLimit40
	laneEnds		speedLimit45
	merge		speedLimit50
	pedestrianCrossing		speedLimit55
	roundAbout		speedLimit60
	signalAhead		speedLimit65
	slow		
	speedBumpsAhead		
	stopAhead		
	thruMergeLeft		
	thruMergeRight		
	turnLeft		
	turnRight		
	yieldAhead		
	warningUrdbl		
noTurn	noLeftAndUTurn		
	noUTurn		
	noLeftTurn		
	noRightTurn		

ACKNOWLEDGMENT

The authors would like to thank their colleagues at the LISA lab for useful discussion and encouragement, especially Eshed Ohn-Bar for his valuable comments. We would also like to thank the reviewers for their thoughtful comments.

REFERENCES

[1] 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.

[2] United Nations, "Vienna Convention on Road Signs and Signals," 1978.

[3] 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.

[4] A. Møgelmose, D. Liu, and M. M. Trivedi, "U.S. Traffic Sign Detection: Remaining Challenges and Tracking," in *Intelligent Transportation Systems (ITSC), 2014 17th International IEEE Conference on*, Oct 2014.

[5] 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.

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

[7] 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.

[8] 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.

[9] G. Wang, G. Ren, Z. Wu, Y. Zhao, and L. Jiang, "A robust, coarse-to-fine traffic sign detection method," in *Neural Networks (IJCNN), The 2013 International Joint Conference on*, Aug 2013, pp. 1–5.

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

[11] R. Benenson, M. Mathias, R. Timofte, and L. Van Gool, "Pedestrian detection at 100 frames per second," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2903–2910.

[12] C. Liu, F. Chang, and Z. Chen, "Rapid Multiclass Traffic Sign Detection in High-Resolution Images," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 6, pp. 2394–2403, Dec 2014.

[13] 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.

[14] 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, 2012, vol. 7476, pp. 478–487.

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

[16] 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.

[17] 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.

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

[19] 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.

[20] ——, "Modular Traffic Sign Recognition applied to on-vehicle real-time visual detection of American and European speed limit signs," vol. 14, 2009.

[21] P. Dollár, "Piotr's Computer Vision Matlab Toolbox (PMT)," http://vision.ucsd.edu/~pdollar/toolbox/doc/index.html.

[22] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics gems IV*. Academic Press Professional, Inc., 1994, pp. 474–485.

[23] P. Dollar, R. Appel, S. Belongie, and P. Perona, "Fast Feature Pyramids for Object Detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 8, pp. 1532–1545, Aug 2014.

[24] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

[25] E. Ohn-Bar and M. Trivedi, "Fast and robust object detection using visual subcategories," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 179–184.

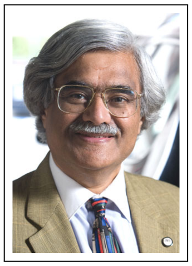
[26] S. Sivaraman and M. Trivedi, "Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 14, no. 4, pp. 1773–1795, Dec 2013.

[27] A. Tawari, A. Møgelmose, S. Martin, T. B. Moeslund, and M. M. Trivedi, "Attention Estimation by Simultaneous Analysis of Viewer and View," in *Intelligent Transportation Systems (ITSC), 2014 17th International IEEE Conference on*, Oct 2014.

[28] D. Shinar, *Traffic safety and human behaviour*. Emerald Group Publishing, 2007.

Andreas Møgelmose received his Bachelor's degree in Computer Engineering on the topic of Information Processing Systems in 2010 and his MSc in Informatics in 2012, both from Aalborg University, Denmark. He is currently pursuing a PhD in the field of visual analysis of people and advanced driver assistance systems, also at Aalborg University. Furthermore, he is a part of the LISA lab at the University of California, San Diego in the Computer Vision and Robotics Research Laboratory, where he has been a visiting scholar multiple times. His main interests are computer vision and machine learning, especially in the area of detecting people, pedestrians, and traffic signs, as well as re-identification of people.

Dongran Liu received his Bachelor of Science in Electrical & Computer Engineering in 2013 from Miami University. He is currently pursuing a Master of Science in Signal and Image Processing from University of California, San Diego. Now, he works in the LISA lab in Computer Vision and Robotics Research Laboratory of UC San Diego. His main interests are machine learning and computer vision.



Mohan Manubhai Trivedi is a Professor of Electrical and Computer Engineering and the Founding Director of the Computer Vision and Robotics Research Laboratory and Laboratory for Intelligent and Safe Automobiles (LISA) at the University of California, San Diego. He and his team are currently pursuing research in machine and human perception, machine learning, human-centered multimodal interfaces, intelligent transportation, driver assistance and active safety systems. He was the Program Chair of the IEEE Intelligent Vehicles Symposium in 2006 and General Chair of the IEEE IV Symposium in 2010. He is an Editor of the IEEE Transactions on Intelligent Transportation Systems and Image and Vision Computing Journal. He is elected to the Board of the ember of the IEEE ITS Society. He served on the Executive Committee of the California Institute for Telecommunication and Information Technologies as the leader of the Intelligent Transportation Layer at UCSD and as a charter member and Vice Chair of the Executive Committee of the University of California System wide Digital Media Initiative. Trivedi serves as a consultant to industry and government agencies in the U.S. and abroad, including the National Academies, major auto manufactures and research initiatives in Asia and Europe. Trivedi is a Fellow of the IEEE ("for contributions to Intelligent Transportation Systems field"), Fellow of the IAPR ("for contributions to vision systems for situational awareness and human-centered vehicle safety"), and Fellow of the SPIE ("for distinguished contributions to the field of optical engineering").