

Vision based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems

Perspectives and Survey

Møgelmoose, Andreas; Trivedi, Mohan M.; Moeslund, Thomas B.

Published in:

I E E E Transactions on Intelligent Transportation Systems

DOI (link to publication from Publisher):

[10.1109/TITS.2012.2209421](https://doi.org/10.1109/TITS.2012.2209421)

Publication date:

2012

Document Version

Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Møgelmoose, A., Trivedi, M. M., & Moeslund, T. B. (2012). Vision based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey. *I E E E Transactions on Intelligent Transportation Systems*, 13(4), 1484-1497. <https://doi.org/10.1109/TITS.2012.2209421>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Vision-Based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey

Andreas Møgelmoose, Mohan Manubhai Trivedi, and Thomas B. Moeslund

Abstract—In this paper, we provide a survey of the traffic sign detection literature, detailing detection systems for traffic sign recognition (TSR) for driver assistance. We separately describe the contributions of recent works to the various stages inherent in traffic sign detection: segmentation, feature extraction, and final sign detection. While TSR is a well-established research area, we highlight open research issues in the literature, including a dearth of use of publicly available image databases and the overrepresentation of European traffic signs. Furthermore, we discuss future directions of TSR research, including the integration of context and localization. We also introduce a new public database containing U.S. traffic signs.

Index Terms—Active safety, human-centered computing, machine learning, machine vision, object detection.

I. INTRODUCTION

IN THIS paper, we provide a survey of traffic sign detection for driver assistance. State-of-the-art research utilizes sophisticated methods in computer vision for traffic sign detection, which has been an active area of research over the past decade. On-road applications of vision have included lane detection, driver distraction detection, and occupant pose inference. As described in [1]–[3], it is crucial to not only consider the car's surrounding and external environment when designing an assist system but also consider the internal environment and take the driver into account. Fusing other types of information with the sign detector, as described in [4], can make the overall system even better.

When the system is considered a distributed system where the driver is an integral part, it allows for the driver to contribute what he is good at (e.g., seeing speed limit signs, as we shall see later), while the TSR part can present information from other signs. In addition, other surround sensors can also have an influence on what is presented.

In recent years, speed limit detection systems have been included in top-of-the-line models from various manufacturers,

TABLE I
SIGNIFICANT RESULTS FROM [9] REGARDING ATTENTION TO VARIOUS SIGN TYPES

Target	Fixated		Not fixated	
	Recalled	Not recalled	Recalled	Not recalled
Speed limit 80 km sign	100	0	0	0
Game Crossing sign	60	0	7	33
Pedestrian crossing ahead	8	54	0	38
Pedestrian crossing sign	0	21	0	79

but a more general sign detection solution and an integration into other vehicle systems have not yet materialized. Current state-of-the-art TSR systems utilize neither information about the driver nor input from the driver to enhance performance. Extensive studies in human-machine interactivity are necessary to present the TSR information in a careful way to inform the driver without causing distraction or confusion. The literature features just two surveys on TSR: In [5], there is a good introduction, but it is not very comprehensive. In [6], any improvements in the field from the past five years are not presented because the paper is several years old. A very good comparison of various segmentation methods is offered in [7], but given that it only covers segmentation, it is not a comprehensive overview of detection methods. Likewise, [8] provides good comparison of Hough transform derivatives. In this paper, our emphasis is on framing the TSR problem in the context of human-centered driver assistance systems. We provide a comparative discussion of papers published mostly within the last five years and an overview of the recent work in the area of sign detection, which is a subset of the TSR problem.

We provide a critical review of traffic sign detection and offer suggestions for future research areas in this challenging problem domain. The next section establishes the driver assistance context and covers TSR systems in general. Section III provides a problem description and a gentle introduction to traffic sign detection. Section IV deals with segmentation for traffic sign detection. Section V details models and feature extraction. Section VI deals with the detection itself. In the final section, analysis and insight on future research directions in the field are provided.

II. HUMAN-CENTERED TRAFFIC SIGN RECOGNITION FOR DRIVER ASSISTANCE: ISSUES AND CONSIDERATIONS

Traffic sign recognition (TSR) research needs to take into account the visual system of the driver. This can include factors

Manuscript received January 30, 2012; revised April 27, 2012; accepted June 19, 2012. Date of publication October 19, 2012; date of current version November 27, 2012. The Associate Editor for this paper was J. Stallkamp.

A. Møgelmoose and T. B. Moeslund are with the Visual Analysis of People Laboratory, Aalborg University, 9220 Aalborg East, Denmark (e-mail: am@create.aau.dk).

M. M. Trivedi is with the Computer Vision and Robotics Research Laboratory, University of California, San Diego, La Jolla, CA 92093-0434 USA.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TITS.2012.2209421

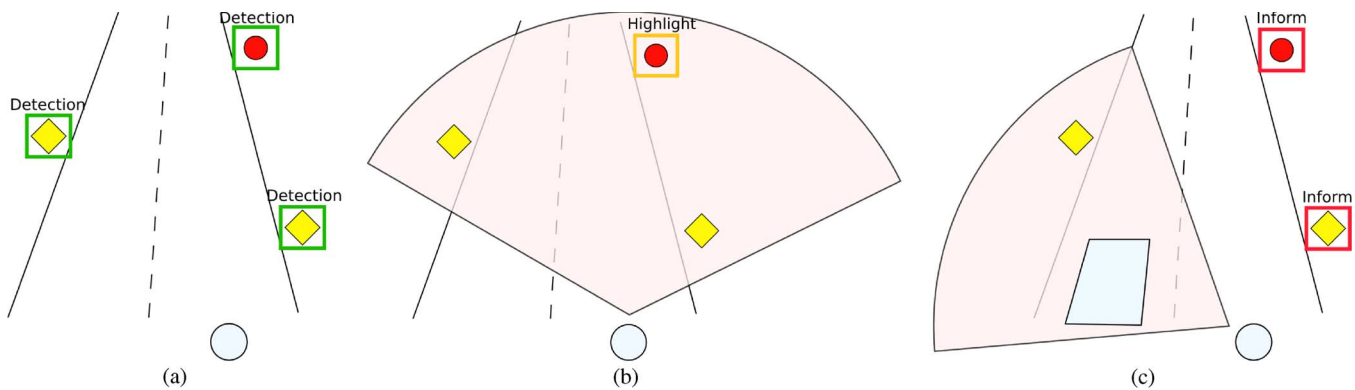


Fig. 1. Different detection scenarios. The circle is the ego car, and three signs are distributed along the road. The area highlighted in red illustrates the driver's area of attention. (a) Standard scenario used for autonomous cars. Here, all signs must be detected and processed. (b) and (c) System that tracks the driver's attention. In (b), the driver is attentive and spots all signs. Therefore, the system just highlights the sign that is known to be difficult for people to notice. In (c), the driver is distracted by a passing car and thus misses two signs. In this case, the system should inform the driver about the two missed signs.

such as visual saliency of signs, driver focus of attention, and cognitive load. According to [9] (see Table I for a summary of the main results), not all signs are equal in their ability to capture the attention of the driver. For example, a driver may fixate his gaze on a sign but neither notice the sign nor remember its informational content. While drivers invariably fixate on speed limit signs and recall their information, they are less likely to notice game crossing and pedestrian signs. This can endanger pedestrians, as it may not leave enough reaction time to stop.

The implications of use of TSR in human-in-the-loop system are clear; instead of focusing on detection and perfectly recognizing all signs of some class, which would be the objective for an autonomous car, the task is now to detect and highlight signs that the driver has not seen. This gives way to various models of TSR, which take into account the driver's focus of attention, and interactivity issues. Driver attention tracking is covered in [10] and [11]. Fig. 1 shows examples on how TSR can be used for driver assistance. Fig. 1(a) shows how a system should act in an autonomous car. It simply recognizes all signs present. In Fig. 1(b), there is a driver in the loop, and while the system may see all the signs, it should avoid presenting them to avoid driver confusion. Instead, it simply highlights the sign type that is easy to overlook, such as the pedestrian crossing warnings in the research. Fig. 1(c) shows how a driver is distracted by a passing car. This causes him to miss two signs. His car has a TSR system for driver assistance, which informs him of the signs as he returns his attention to the road ahead of him. This could, for example, be done using a heads-up display, as suggested in [12].

Even though this paper is mostly concerned with using TSR for driver assistance, TSR has various well-defined applications nicely summarized here [13].

- 1) Highway maintenance: Check the presence and condition of signs along major roads.
- 2) Sign inventory: Similar to the preceding task, create an inventory of signs in city environments.
- 3) Driver-assistance systems: Assist the driver by informing of current restrictions, limits, and warnings.
- 4) Intelligent autonomous vehicles: Any autonomous car that is to drive on public roads must have a means of

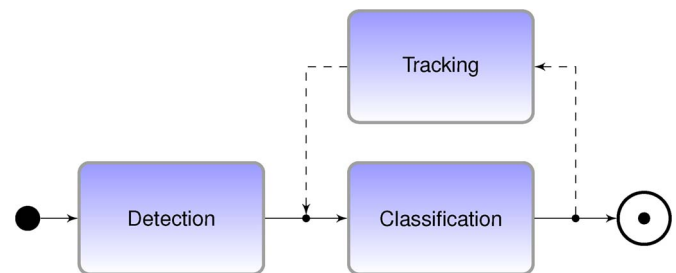


Fig. 2. Basic flow in most TSR systems.

obtaining the current traffic regulations. This can be done through TSR.

This paper uses the term TSR to refer to the entire chain from detection of signs to their classification and potential presentation to the driver. Generally, TSR is split into two stages: detection and classification (see Fig. 2). Detection is concerned with locating signs in input images, whereas classification is about determining what type of sign the system is looking at. The two tasks can often be treated as completely separate, but in some cases, the classifier relies on the detector to supply information, such as the sign shape or sign size. In a full system, the two stages depend on each other, and it does not make sense to have a classifier without a detection stage. Later, we divide the detection stage into three substages, but these should not be confused with the two main stages of a full TSR system: detection and classification.

Apart from shape and color, another aspect may be used in TSR: temporal information. Most TSR systems are designed with a video feed from a vehicle in mind; therefore, signs can be tracked over time. The simplest way of using tracking is to accept sign candidates as signs only if they have shown up on a number of consecutive frames. Sign candidates that only show up once are usually a result of noise. Employing a predictive method, such as a Kalman filter, allows for the system to predict where a sign candidate should show up in the next frame, and if its position is too far away from this prediction, the sign candidate is discarded. A predictive tracking system has the additional benefit of handling occlusions, hence, preventing signs that were occluded from being classified as new signs. This is very important in a driver-assistance system where

signs should only be presented once and in a consistent way. Imagine a scenario where a sign is detected in a few frames and occluded for a short time before being detected again. For an autonomous car, it is not likely to be a problem to be presented with the same information twice: If the first sign prompted the speed to be set at 55 mi/h, there is no problem in the system being told once again that the speed limit is 55 mi/h. In a driver-assistance system, the system must not present more information than absolutely necessary at any given moment, so the driver is not overwhelmed with information, e.g., forcing the driver to pay attention to a sign he has already seen should be avoided.

Many TSR systems are tailored to a specific sign type. Due to the vast differences in sign design from region to region (see the next section) and the differences in sign design based on their purpose, many systems narrow their scope down to a specific sign type in a specific country.

There is a wide span in speeds of the systems. For use in driver assistance and autonomous vehicles, real-time performance is necessary. This does not necessarily mean a speed of 30 Hz, but the signs must be read quickly enough to still be relevant to act on. Depending on the exact application, a few hertz is required.

Instead of treating the entire TSR process in what could easily become a cursory manner, we have opted to thoroughly look on the detection stage. The line between detection and classification is a bit blurry since some detectors provide more information to the classifier than others. It is normal for the detector to inform the classifier of the general category of signs since that is often defined by either the overall sign shape or its color, which is something that the detector itself may use to localize the sign.

Even though this paper is targeted toward the problem of detecting traffic signs, one must not forget that, without a subsequent classification stage, the systems are useless. Thus, even though we encourage a decoupling of the two tasks, this does not mean that the classification is a solved problem. It is a crucial part of a full system.

III. TRAFFIC SIGNS

Traffic signs are markers placed along roads to inform drivers about either road conditions and restrictions or which direction to go. They communicate a wealth of information but are designed to do so efficiently and at a glance. This also means that they are often designed to stand out from their surroundings, making the detection task fairly well defined.

The designs of traffic signs are standardized through laws but differ across the world. In Europe, many signs are standardized via the Vienna Convention on Road Signs and Signals [14]. There, shapes are used to categorize different types of signs: Circular signs are prohibitions including speed limits, triangular signs are warnings, and rectangular signs are used for recommendations or subsigns in conjunction with one of the other shapes. In addition to these, octagonal signs are used to signal a full stop, and downward-pointing triangles signal a yield. Countries have other different types, e.g., to inform about city limits. Examples of these signs can be seen in Fig. 3.

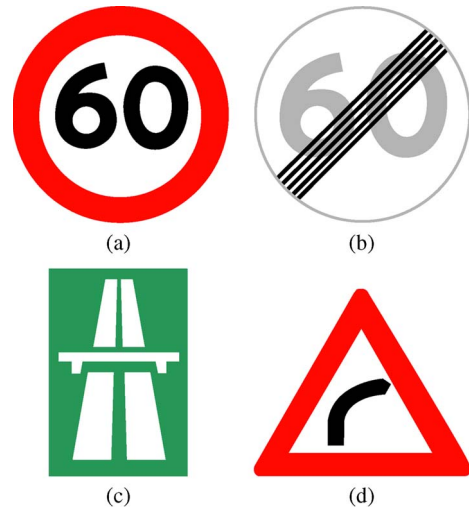


Fig. 3. Examples of European signs. These are Danish, but many countries use similar signs. (a) Speed limit. Sign C55. (b) End speed limit. Sign C56. (c) Start of freeway. Sign E55. (d) Right turn. Sign A41.

In the U.S., traffic signs are regulated by the Manual on Uniform Traffic Control Devices (MUTCD) [15]. It defines which signs exist and how they should be used. It is accompanied by the Standard Highway Signs and Markings (SHSM) book, which describes the exact designs and measurements of signs. At the time of writing, the most recent MUTCD was from 2009, whereas the SHSM book had not been updated since 2004. Thus, it described the MUTCD from 2003. The MUTCD contains a few hundred different signs divided into 13 categories.

To further complicate matters, each U.S. state can decide whether it wishes to follow the MUTCD. A state has three options.

- 1) Adopt the MUTCD fully as is.
- 2) Adopt the MUTCD but add a State Supplement.
- 3) Adopt a State MUTCD that is “in substantial conformance with” the national MUTCD.

In the U.S., 19 states have adopted the national MUTCD without modifications, 23 states have adopted the national MUTCD with a state supplement, and ten states have opted to create a State MUTCD (the count includes the District of Columbia and Puerto Rico). Examples of U.S. signs can be seen in Fig. 4.

New Zealand uses a sign standard with warning signs that are yellow diamonds, as in the U.S., but regulatory signs are round with a red border, like those from the Vienna Convention countries. Japan uses signs that are generally in compliance with the Vienna Convention, as are Chinese regulatory signs. Chinese warning signs are triangular with a black/yellow color scheme. Central and South American countries do not participate in any international standard but often use signs somewhat like the American standard.

While signs are well defined through laws and designed to be easy to spot, there are still plenty of challenges for TSR systems.

- 1) Signs are similar within or across categories (see Fig. 5).
- 2) Signs may have faded or are dirty so they are no longer their specified color.

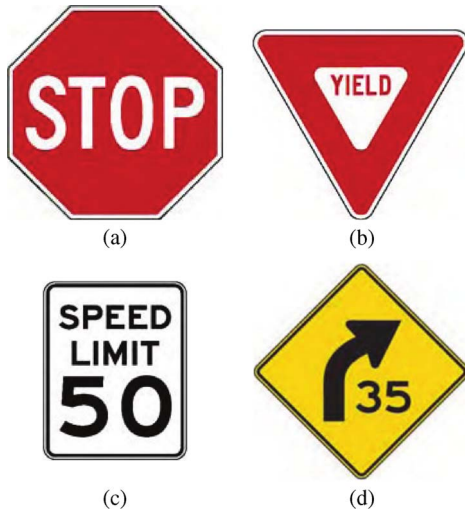


Fig. 4. Examples of signs from the U.S. national MUTCD. Image source: [15]. (a) Stop. Sign R1-1. (b) Yield. Sign R1-2. (c) Speed limit. Sign R2-1. (d) Turn warning with speed recommendation. Sign W1-2a.

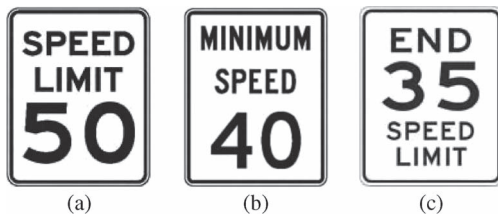


Fig. 5. Examples of similar signs from the MUTCD. The situation in (c) exists only in the California MUTCD. Image source: [15]. (a) Speed limit. Sign R2-1. (b) Minimum speed. Sign R2-4. (c) End speed limit. Sign R3 (CA).

- 3) The sign post may be bent, and therefore, the sign is no longer orthogonal to the road.
- 4) Lighting conditions may make color detection unreliable.
- 5) Low contrast may make shape detection hard.
- 6) In cluttered urban environments, other objects may look very similar to signs.
- 7) There may be varying weather conditions.

A. Assessing Performance of Sign Detectors

When comparing sign detectors, some comparison metrics must be set up. The straightforward and most important measure is the true positive rate. However, even if all signs are detected, the system is not necessarily perfect. The number of false positives must also be taken into account. If the amount of false positives is too high, the classifier will have to handle a lot more data than it should, degrading the overall system speed. For cases when a system must work in real time in a car, obviously, the detection must be fast. In general, the faster the detection runs, the more time left over for the classification stage. Adjusting these goals is a tradeoff. Often, the target will be to create a system that is just fast enough for a given application while keeping the receiver operating characteristic acceptable. Another interesting performance characteristic is which sign types a given system works for.

Even with the parameters in mind and a clear idea of the performance metrics, comparing the performance of different systems is not a straightforward task. Unlike other computer vision areas, until recently, no standardized training and test data set existed, so no two systems were tested with the same data. The image quality varies from high-resolution still images (as in [16]–[18]) to low-resolution frames from in-car video cameras (such as [19]–[21]). That, combined with the facts that signs wildly vary between countries and many papers limit their scope to specific sign types, makes for a quite uneven playing field.

For a discussion of the performance of the papers presented in this survey, see Section IV.

B. Public Sign Databases

A few publicly available traffic sign data sets exist:

- 1) German TSR Benchmark (GTSRB) [22], [23];
- 2) KUL Belgium Traffic Signs Data set (KUL Data set) [24];
- 3) Swedish Traffic Signs Data set (STS Data set) [25];
- 4) RUG Traffic Sign Image Database (RUG Data set) [26];
- 5) Stereopolis Database [27].

Information on these databases can be found in Table II. Most of the databases have emerged within the last two years (except for the very small RUG Data set) and are not yet widely used. One of the most widespread databases is the GTSRB, which has been presented in [22] and created for the competition “The German Traffic Sign Recognition Benchmark.” The competition was held at the International Joint Conference on Neural Networks (IJCNN) 2011. It is a large data set containing German signs and is thus very suitable for training and testing systems aimed at signs adhering to the Vienna Convention. A sample image from the GTSRB database can be found in Fig. 6(a). The GTSRB is primarily geared toward classification, rather than detection, since each image contains exactly one sign without much background. For detection, images of complete scenes are necessary. In addition, many detection systems rely on a tracking scheme to make detection more robust, and without video of the tracks (in GTSRB parlance, a “track” is a set of images of the same physical sign), this will not properly work. Since the data set is created for the classification task, this is not so much a problem of that database, as it is a testament to its target. In conjunction with the competition, five interesting papers [28]–[32] were released. They all focus on classification rather than detection.

Two other data sets should be highlighted: The STS Data set and the KUL Data set. They are both very large, although not as large as the GTSRB, and they contain full images. This means that they can both be used for detection purposes. The STS Data set does not have all images annotated, but it does include all frames from the videos used to obtain the data. This means that tracking systems can be used on this data set, but it can only be verified with ground truth every five frames. An example from the STS Data set can be seen in Fig. 6(b). The KUL Data set also includes four recorded sequences, which can be used for tracking experiments. KUL also includes a set of sign-free images, which can be used as negative training images, and it has pose information for the cameras for each image.

TABLE II
INFORMATION ON THE PUBLICLY AVAILABLE SIGN DATABASES

	GTSRB	STS Dataset	KUL Dataset	RUG Dataset	Stereopolis	LISA Dataset
Number of classes:	43	7	100+	3	10	49
Number of annotations:	50000+	3488	13444	0	251	7855
Number of images:	50000+	20000	9006	48	847	6610
Annotated images:	All images	4000 images	All images	0	All images	All images
Sign sizes:	15x15 to 250x250 px	3x5 to 263x248 px	100x100 to 1628x1236 px	N/A	25x25 to 204x159 px	6x6 to 167x168 px
Image sizes:	15x15 to 250x250 px	1280x960 px	1628x1236 px	360x270 px	1920x1080 px	640x480 to 1024x522 px
Includes videos:	No	No	Yes, 4 tracks	No	No	Yes, for all annotations
Country of origin:	Germany	Sweden	Belgium	The Netherlands	France	United States
Extra info:	Images come in tracks with 30 different images of the same physical sign.	Signs marked visible/blurred/occluded and whether they belong to the current road or a side road.	Includes traffic sign annotations, camera calibrations and poses.	Does not include any annotations, only raw pictures.		Images from various camera types.

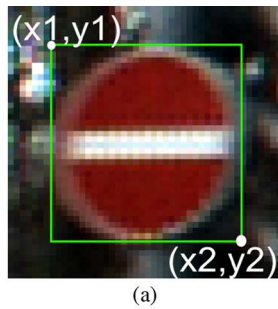


Fig. 6. Example sign images from (a) the GTSRB and (b) the STS Data set, with the sign bounding boxes superimposed.

From the research, it was evident that there was a lack of databases with U.S. traffic signs, and therefore, in conjunction with this paper, we have assembled one. Its details are also listed in Table II. One novel feature of this data set is that it includes video tracks of all the annotated signs. Many systems already use various tracking schemes to minimize the number

of false positives, and it is quite likely that, in the future, detectors using temporal data will emerge even more. Therefore, the LISA data set includes video and standalone frames. Not all frames have been extracted for annotation, but all annotated frames can be traced back to the source video; therefore, so the annotations can also be used to verify systems using tracking.

IV. SIGN DETECTION

The approaches in this stage have traditionally been divided into two kinds:

- 1) color-based methods;
- 2) shape-based methods.

Color-based methods take advantage of the fact that traffic signs are designed to be easily distinguished from their surroundings, often colored in highly visible contrasting colors. These colors are extracted from the input image and used as a base for the detection. Just as signs have specific colors, they also have very well-defined shapes that can be searched for. Shape-based methods ignore the color in favor of the characteristic shape of signs.

Each method has its pros and cons. Color of signs, while well defined in theory, varies much with available lighting, as well as with the age and condition of the sign. On the other hand, searching for specific colors in an image is fairly straightforward. Sign shapes are invariant to lighting and age, but parts of the sign can be occluded, making the detection harder or the sign may be located at a background of a similar color, ruining edge detection, on which most shape detectors rely.

The division of systems in this way can be problematic. Almost all color-based approaches take shape into account after having looked at colors. Others use shape detection as their main method but integrate some color aspects as well. Instead, the detection can be split into two steps, as proposed in [7], i.e., segmentation and detection. In this paper, we go one step further and split the detection step into a feature extraction step

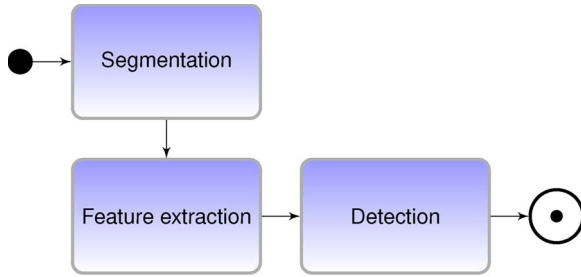


Fig. 7. General flow followed by typical sign detection algorithms.

and the actual detection, which acts on the features that are extracted. Many shape-only-based methods have no segmentation step. The flow is outlined in Fig. 7.

An overview of all surveyed papers and their methods is listed in Table III. It contains each of the systems and lists which segmentation method, feature type, and detection method are used. The author group numbers are used to mark the papers that are part of an ongoing effort from the same group of authors. They do not constitute a ranking in any way. In Tables IV and V, some of their more detailed properties are listed. The systems are split into two tables. Table IV displays those that do not use any tracking. Table V contain those that do use tracking, something we find crucial when using TSR in a driver-assistance context, as mentioned earlier. Apart from this division, the two tables are structured in the same way: *Sign type in paper* describes which sign types the authors of the paper have attempted to find, whereas *sign type possible* are the types of signs the method could be extended to include, which is usually a very broad group. *Real time* is about how fast the system runs, if that information is available. Any system with a frame rate faster than 5 frame/s is considered to have real-time potential. *Rotation invariance* tells whether the used technique is robust to rotation of signs. *Model versus training* describes if the detection system relies on a theoretical model of signs (such as a predefined shape), if it uses a learned type of classifier, or if it uses a combination of the two. *Test image type* is the image resolution that the system is designed to work with. Low-resolution images are usually video frames, whereas high-resolution images are still images.

The detection performance of the surveyed papers is presented in Table VI. As mentioned earlier, very few papers use common databases to test their performance, and the papers detect various types and numbers of signs. Thus, the numbers should not be directly compared; nevertheless, they give an idea of performance. Not all papers report all the measures reported in the table (detection rate, false positives per frame, etc.), so some fields in the table could not be filled. In other cases, these exact measures were not given but could be calculated from other given numbers. Where figures are available, the best detection rate that the system obtained is reported, along with the corresponding measure of false positives. The detection rate is per frame, meaning that 100% detection is only achieved if a sign is found in every frame present. It is not sufficient to just detect the sign in a few frames. This is the way results are presented in most papers, and therefore, this is the measure chosen here, even if a real-world system works well enough

if each sign is just detected once. Papers that only report the per-sign detection rate as opposed to the per-frame detection rate are marked with a triangle in the rightmost column of the table.

Different papers report the false positives in different ways, so a few different measures, which are not directly comparable, are presented in the table:

FPPF) False positives per frame: $FPPF = FP/f$, where FP is the number of false positives, and f is the number of frames analyzed.

FPR) False positive rate: $FPR = FP/N$, where N is the number of negatives in the test set. This measure is rarely used in detection since the number of negatives does not always make much sense (how many negatives exist in a full frame).

PPV) Positive predictive value: $PPV = TP/TP + FP$, where TP is the number of true positives.

FPTP) False/true positive ratio: $FPTP = FP/TP$.

WPA) Wrong pixels per area: $WPA = WP/AP$, where WP is the number of wrongly classified pixels, and AP is the total number of pixels classified.

When papers present results for different sign types, the mean detection performance is also presented in the table. In many cases, that will give a better view of the true performance of the approach.

Five papers stick out, claiming 100% detection rate. The first [33] is only tested on synthetic data. It is possible that the synthetic data do not fully encapsulate real-world variations, so the performance of that approach is not guaranteed to be as good in real-world scenarios. At first glance, [34] achieves 100% detection rate, but that is only the case for one of their sign types. The mean performance is a more accurate (and still promising) gauge of the actual performance. The same is the case for [25]. In [35], all signs in the test set are detected, but at the cost of a large number of false positives per frame. In [36], the per-sign detection rate is all that is presented, and therefore, the figure cannot be compared with other systems.

Generally, systems achieve detection rates well into the 90% range, whereas some achieve very low false detection rates. From the table, no “best system” can be chosen since the test sets are very different in both size and content. A system that can detect several different sign types at low detection rate may, in some applications, be considered better than a system that can only detect one specific sign type but does that very well. A few papers that should be highlighted are [18] and [37]–[39]. They have all been tested on large data sets and report detection rates above 90% with a decent low number of false positives.

Now that the basics about sign detection are in place, the succeeding sections go in depth with how recent papers perform each step.

V. SEGMENTATION

The purpose of the segmentation step is to achieve a rough idea about where signs might be and thus narrow down the search space for the next steps. Not all authors make use of this step. Since the segmentation is traditionally done based on colors, authors who believe this should not be part of sign

TABLE III
OVERVIEW OF DETECTION METHODS IN 41 RECENT PAPERS. PAPERS WITH THE SAME BACKGROUND COLOR ARE PAPERS WRITTEN BY THE SAME GROUP. WHITE BACKGROUND INDICATE STANDALONE PAPERS

Paper	Year	Author group	Segmentation method	Features	Detection method
[16]	2005	1	HSI thresholding with addition for white signs ([49])	Boundary distance transform	Correlation with model distance transforms
[47]	2007	1	HSI thresholding with addition for white signs ([49])	DtB (distance to bounding box)	Linear SVM
[33]	2008	1	HSI thresholding with addition for white signs ([49])	FFT of shape signatures	Euclidian nearest neighbor
[48]	2010	1	HSI thresholding with addition for white signs ([49])	DtB (distance to bounding box)	Linear SVM
[66]	2005	2	None	Haar wavelet features computed on specific color channels	Cascaded classifier
[40]	2008	2	Extended radial symmetry voting	Haar wavelet features	Cascaded classifier
[17]	2006	3	LCH thresholding (obtained with CIECAM97)	HOG	Comparison with template vectors
[50]	2008	3	LCH thresholding (obtained with CIECAM97)	None	None
[62]	2007	4	None	Color filtered edges	Extended radial symmetry voting
[34]	2010	4	HSV discretization	Edges	Extended radial symmetry voting
[20]	2011	4	Quad-tree color selection	Edges	Extended radial symmetry voting
[35]	2004	5	None	Edges	Extended radial symmetry voting
[54]	2006	5	None	Edges	Extended radial symmetry voting
[55]	2008	5	None	Edges	Radial symmetry voting
[56]	2008	6	None	Edges	Votes for symmetric areas to be used as ROI with another shape-detector
[57]	2011	6	None	Edges	Two-tier radial symmetry voting
[58]	2011	7	None	Edges of closed contours with certain aspect ratios	Hough shape detection
[59]	2011	7	None	Edges of closed contours with certain aspect ratios	Hough shape detection
[18]	2009	8	Adaptive RGB threshold	Edges	Fuzzy templates (a Hough derivative)
[51]	2010	8	Adaptive RGB threshold	Edges and Haar-like features	Fuzzy templates, cascaded classifier, and SVM
[64]	2008	9	None	HistFeat (HOG derived)	Cascaded classifier
[39]	2011	9	None	Various HOG-features	5 stage cascaded classifier trained with LogitBoost
[49]	2002	None	HSI thresholding with edge detection and removal of achromatic colors	Edges	Genetic algorithm looking for circles
[63]	2007	None	None	Edge orientation histograms	Comparison with template vectors
[41]	2007	None	HSI thresholding	Edges	Hough shape detection
[60]	2007	None	None	Edges	Hough transform for circular signs, proprietary (not described) for rectangular
[42]	2008	None	HSI thresholding	30x30 px YcbCr patches	Neural network
[37]	2009	None	None	Dissociated dipoles	Cascaded classifier
[61]	2009	None	None	Edges	Vertex and Bisector transform (VBT)
[36]	2009	None	Radial symmetry voting combined with SIFT features	Edge colors	Contracting Curve Density
[43]	2009	None	HSV thresholding	Edges	Hough shape detection
[19]	2009	None	Saliency detection with color and edges	HOG	SVM
[44]	2009	None	Hue thresholding on chromatic colors only	Tangent function of simplified contours	Distance from model tangent function
[45]	2010	None	HSI thresholding	Edges	Circle center voting
[65]	2010	None	None	HOG augmented with color information	SVM
[21]	2010	None	Biologically inspired attention model	Color, corner positions, height, excentricity	Color, corner positions, height, excentricity
[46]	2010	None	HSI thresholding	Edges	Radial symmetry voting
[52]	2011	None	Nested cascade classifier with Local Rank Pattern features (based on 7 RGB based colors)	Edges	RANSAC circle fit
[38]	2011	None	Radial symmetry voting	Colors in modified RGB-space	Distance to learned colors
[8]	2011	None	Probabilistic color preprocessing	Edges	Hough derivative shape detector
[25]	2011	None	None	Fourier descriptors	Correlation based matching

TABLE IV
OVERVIEW OF DETAILED PROPERTIES OF THE 27 PAPERS THAT DO NOT USED TRACKING

Paper	Year	Author group	Sign type in paper	Sign type possible	Real-time	Rotation inv.	Model vs. training	Test image type
[16]	2005	1	5 regular polygons, various colors	Colored	N/A	Yes	Both	High-res
[33]	2008	1	Circular, triangular, square and semiellipses	Colored	N/A	Yes	Both	Low-res
[48]	2010	1	Circular red	Circular colored	No	No	Both	Low-res
[17]	2006	3	Circular red, circular blue and triangular red	Colored	No	Yes	Both	High-res
[50]	2008	3	Circular red and blue	Colored	N/A	Yes	Training	High-res
[62]	2007	4	Circular, triangular and square, various colors	Regular, colored polygons	Yes	No	Both	N/A
[35]	2004	5	Regular polygons	Regular polygons	Yes	Yes	Model	Low-res
[54]	2006	5	Regular polygons	Regular polygons	Yes	No	Model	Low-res
[55]	2008	5	Circular red	Circular	Yes	Yes	Model	Low-res
[18]	2009	8	Circular red, circular blue, diamond white	Colored	No	No	Both	High-res
[64]	2008	9	Circular, triangular and octagonal red	Any sign	Yes	No	Training	Low-res
[39]	2011	9	Circular red	Any sign	Yes	No	Training	Low-res
[49]	2002	None	Circular, red	Colored	N/A	Yes	N/A	N/A
[63]	2007	None	Circular, triangular, diamond, octagonal	Any sign	Yes	No	Training	Low-res
[41]	2007	None	Circular and triangular, red	Colored	N/A	Yes	N/A	N/A
[42]	2008	None	Circular red, triangular red, and octagonal red	Colored	No	No	Training	N/A
[37]	2009	None	Circular and triangular, red	Any sign	No	No	Training	N/A
[61]	2009	None	Triangular red and blue	Triangular	Yes	Yes	Model	Low-res
[43]	2009	None	Circular, triangular, square, various colors	Colored	No	Yes	Model	Low-res
[19]	2009	None	Circular, red and square blue	Colored	N/A	Yes	Training	Low-res
[44]	2009	None	Circular, triangular, square, various colors	Colored	N/A	No	Both	Low-res
[45]	2010	None	Circular red	Colored	No	No	Model	Low-res
[65]	2010	None	Circular red, circular blue and triangular red	Any sign	N/A	No	Training	High-res
[21]	2010	None	Triangular and octagonal, red	Any sign	Yes	N/A	Model	N/A
[46]	2010	None	Circular red	Colored	N/A	No	Both	Low-res
[8]	2011	None	Circular and triangular, red and blue	Regular, colored polygons	No	No	Both	High-res
[25]	2011	None	7 sign types	Any sign	N/A	Yes	Model	Low-res

No

TABLE V
OVERVIEW OF DETAILED PROPERTIES OF THE 14 PAPERS THAT USED TRACKING

Paper	Year	Author group	Sign type in paper	Sign type possible	Real-time	Rotation inv.	Model vs. training	Test image type
[47]	2007	1	Circular and triangular, red	Colored	No	No	Both	Low-res
[66]	2005	2	Circular red	Any sign	Yes	No	Training	Low-res
[40]	2008	2	Rectangular white	Any sign	Yes	No	Both	Low-res
[34]	2010	4	40 different signs	Any sign	Yes	No	Model	Low-res
[20]	2011	4	Circular red and blue	Regular, colored polygons	Yes	No	Both	Low-res
[56]	2008	6	Circular, triangular and octagonal	Any sign	Yes	No	Model	N/A
[57]	2011	6	Circular	Circular	Yes	Yes	Model	Low-res
[58]	2011	7	100 different signs, circular and triangular	Any sign	Yes	Yes	Model	Low-res
[59]	2011	7	8 sign categories	Any sign	No	Yes	Model	Low-res
[51]	2010	8	N/A	Colored	Yes	N/A	Both	Low-res
[60]	2007	None	Circular red and rectangular white	Circular and rectangular	N/A	N/A	Model	Low-res
[36]	2009	None	Circular red	Circular	N/A	Yes	Both	Low-res
[52]	2011	None	Circular red	Colored	No	Yes	Training	Low-res
[38]	2011	None	Circular red and blue	Circular colored	Yes	Yes	Both	Low-res

TABLE VI
OVERVIEW OF THE PERFORMANCE OF THE PAPERS INCLUDED IN THIS SURVEY. FOR THOSE PAPERS WHERE THE NUMBERS ARE AVAILABLE, THE BEST AND MEAN DETECTION RATES ARE PRESENTED, ALONG WITH THE CORRESPONDING FALSE POSITIVE MEASURE. NOTE THAT THE SYSTEMS HAVE ALL BEEN TESTED IN DIFFERENT WAYS. THEREFORE, A DIRECT COMPARISON IS NOT FEASIBLE (SEE SECTION IV FOR FURTHER DETAILS)

Paper	Year	Group	Evaluation data format	Pos/neg in evaluation data	Best detection rate	False positives for best detection	Mean detection rate	Mean false positives
[16]	2005	1		No statistical results given				
[47]	2007	1	5176 images from 5 videos containing 104 signs	N/A	N/A	N/A	N/A	N/A
[33]	2008	1	2000 synthetic images	N/A	100%	WPA: 0.74%	89.08%	WPA: 13.17%
[48]	2010	1		No statistical results given				
[66]	2005	2	Images from videos	1700 pos/40000 neg	98.6%	FPR: 0.03%	-	-
[40]	2008	2	16828 images from videos	80 positives	98.75%	FPPF: 0.062	-	-
[17]	2006	3	Images from videos	98 positives	95%	N/A	-	-
[50]	2008	3	128 images	142 positives	94%	PPV: 23%	89.67%	PPV: 26%
[62]	2007	4		No results given for the detection stage only				
[34]	2010	4	Images from videos containing 210 signs	N/A	100%	N/A	92.9%	N/A
[20]	2011	4		No statistical results given (graphs are available in the paper)				
[35]	2004	5	45 images	49 positives	100%	FPPF: 0.67	96.67%	FPPF: 0.56
[54]	2006	5	47 images from 1 video	47 positives	93.62%	FPPF: 2.26	-	-
[55]	2008	5	Images from videos	N/A	93%	FPPF: 0.5	-	-
[56]	2008	6		No statistical results given (graphs are available in the paper)				
[57]	2011	6	Images from 34 videos containing more than 100 signs	N/A	87.12%	FPR: 0.14%	-	-
[58]	2011	7	30000 images from 1 video	340 positives	97.74%	FPPF: 0.0024	96.45%	FPPF: 0.0014
[59]	2011	7	Images from videos containing 500 signs	N/A	99.96%	N/A	99.52%	N/A
[18]	2009	8	7356 images containing 269 signs	2459 positives	95.7%	FPPF: 2.5	-	-
[51]	2010	8		No statistical results given (see [18] instead)				
[64]	2008	9		No statistical results given (graphs are available in the paper)				
[39]	2011	9	Images from videos	21500 pos/40000 neg	98.68%	FPR: $10^{-8}\%$	-	-
[49]	2002	None		No statistical results given				
[63]	2007	None	Video tracks, 10-200 frames in length	105 positives	81.9%	N/A	-	-
[41]	2007	None		No results given for the detection stage only				
[60]	2007	None	Images from videos containing 281 signs	N/A	88.97%	FPPF: 0	-	-
[42]	2008	None	164 images	164 positives	92.45%	N/A	-	-
[37]	2009	None	4755 images from 4 videos	N/A	97%	FPPF: 0.056	92%	FPPF: 0.048
[61]	2009	None	48 images	40 positives	82.5%	FPPF: 0.042	-	-
[36]	2009	None	Images from 1 30 min. video containing 94 signs	N/A	100%	N/A	-	-
[43]	2009	None	Images from videos containing 20 signs	N/A	N/A	N/A	-	-
[19]	2009	None	More than 500 images from videos	N/A	99.16%	FPR: 5.56%	98.3%	FPR: 4.72%
[44]	2009	None	1000 images	N/A	95%	FPTP: 0%	91.8%	FPTP: 0.9%
[45]	2010	None	Images from videos	397 pos/697 neg	89.42%	FPR: 0.05%	-	-
[65]	2010	None	3000 images	N/A	85%	N/A	72.47%	N/A
[21]	2010	None	820 images from 2 videos	117 positives	89.8%	PPV: 98.3%	-	-
[46]	2010	None	85 images from video	95 positives	76.64%	FPPF: 0.094	-	-
[52]	2011	None	2967 images from videos	4886 positives	90.1%	PPV: 85.6%	-	-
[38]	2011	None	2134 images from videos	3298 positives	94.03%	FPPF: 3.41	-	-
[8]	2011	None		Comparison of different methods, thus no final result to report.				
[25]	2011	None	STS dataset	641 positives	95.33%	PPV: 100%	77.08%	PPV: 91.85%

detection often have no segmentation step but directly go to the detection.

Of the papers that do use segmentation, all, except [38] and [40], use colors to some extent. Normally, segmentation is done with colors, and subsequently, shape detection is run in a later stage. In [38], the usual order is reversed; therefore, they use radial symmetry voting (see Section VII) for segmentation and a color-based approach for the detection. In [40], radial symmetry voting as preprocessing is also run, but it was followed up with a cascaded classifier using Haar wavelets (see Section VII).

Generally, color-based segmentation relies on a thresholding of the input image in some color space. Since many believe that the RGB color space is very fragile with regard to changes in lighting, these methods are spearheaded by the hue, saturation, and intensity (HIS) space (or its close sibling, the hue, saturation, and value (HSV) space). HSI/HSV is used by [41]–[46]. The HIS space models human vision better than RGB and allows some variation in the lighting, most notably in the intensity of light. Some papers, like those in the series starting with [16] and followed by [33], [47], and [48], augment



Fig. 8. Example of thresholding, looking for red hues. (a) Before thresholding. (b) After thresholding.

the HSI thresholding with a way to find white signs. Hue and saturation are not reliable for detecting white since it can be at any hue they use an achromatic decomposition of the image proposed by [49] (see Fig. 8).

Some authors are not satisfied with the performance of HSI since it does not model the change in color temperature in different weather but only helps in changing light intensity. References [17] and [50] instead threshold in the luminosity, chroma, and hue (LCH) color space, which is obtained using the CIECAM97 model. This allows them to take variations in color temperature into account. The RGB space is used by [18] and [51], but they use an adaptive threshold in an attempt to combat instabilities caused by lighting variations.

Of special interest in this color space discussion is the excellent paper [7], which has shown that HSI-based segmentation offers no significant benefit over normalized RGB and that methods that use color segmentation generally perform much better than shape-only methods. They do, however, have trouble with white signs. For a long time, it has simply been assumed that the RGB color space was a bad choice for segmentation, but through rigorous testing, they show that there is nothing to gain from switching to the HSI color space, instead of a normalized RGB space. As the authors write: “Why use a nonlinear and complex transformation if a simple normalization is good enough?”

A color-based model not relying on thresholding was put forward in [52], which uses a cascaded classifier trained with AdaBoost, which is similar to that proposed by [53] but on Local Rank Pattern features, instead of Haar wavelets. In addition, [34] used a color-based search method that, while closely related to, is not directly thresholding based. Here, the image is discretized into colors that may exist on signs. The discretization process is less destructive than thresholding in that it does not directly discard pixels; instead, it maps them into the closest sign-relevant color. In a more recent contribution [20], they replace the color discretization method with a Quad-tree interest-region-finding algorithm, which finds interesting areas using an iterative search method for colored signs. In the same realm lies [8], which uses learned probabilistic color preprocessing.

In [21], a unique approach is proposed: Using a biologically inspired attention system, it produces a heat map that denotes areas where signs are likely to be found. An example can be seen in Fig. 9. A somewhat similar system was put forth by [19], who uses a saliency measure to find possible areas of interests.

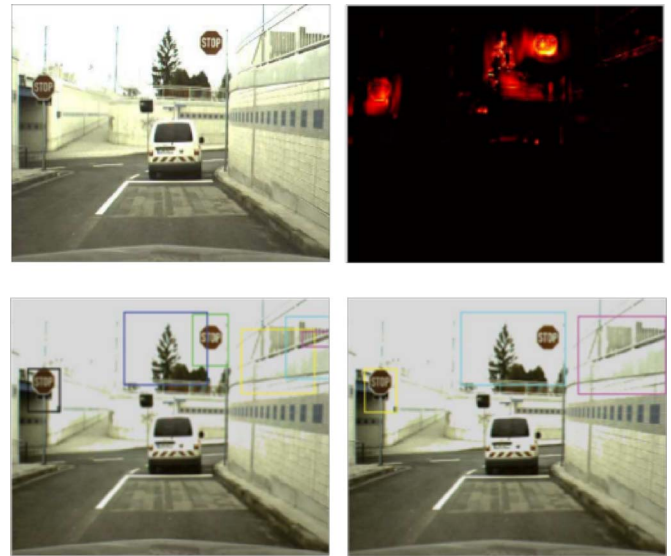


Fig. 9. Biologically inspired detection stage from [21]. Image source: [21].

VI. FEATURES AND MODELING

While various features are available from the vision literature, the choice of feature set is often closely coupled with the detection method, although some feature sets can be used with a selection of different detection methods. The most popular feature is edges: sometimes edges directly obtained from the raw picture and sometimes edges from presegmented images. Edges are practically always found using Canny edge detection or some method that is very similar, and they are used as the only feature in [8], [18], [20], [34], [35], [41], [43], [45], [46], [49], [52], and [54]–[61]. Edges with Haar-like features are combined in [51], and [36] and [62] looked only at certain color-filtered edges.

Even though edges comprise the most popular feature choice, there are other options. Histogram of Oriented Gradients (HOG) is one. It was first used to detect people in images but has been used in [17], [19], [39], [63], and [64] to detect signs. HOG is based on creating histograms of gradient orientations on patches of the image and comparing them to known histograms for the sought-after objects. HOG is also used by Creusen *et al.* [65], but they augment the HOG feature vectors with color information to make them even more robust.

A number of papers [37], [40], [51], [66] use Haar wavelet-like features only on certain colors [66] and in the form of

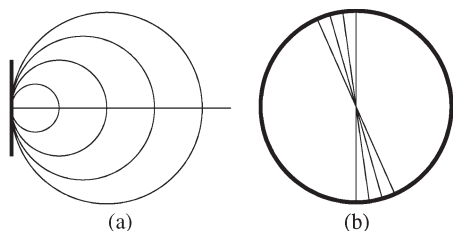


Fig. 10. Basic principle behind the radial symmetry detector. Image inspired by [55]. (a) Possible circles for a gradient. (b) Intersecting vote lines.

so-called dissociated dipoles with wider structure options than traditional Haar wavelets [37].

More esoteric choices are distance to bounding box (DtB), fast Fourier transform (FFT) of shape signatures, tangent functions, simple image patches, and combinations of various simple features. DtB, as used in [47] and [48], are a measure of distances from the contour of a sign candidate to its bounding box. Similarly, the FFT of shape signatures used in [33] is based on the distance from the shape center to its contour at different angles. Tangent functions, which are used in [44], calculate the angles of the tangents at various points around the contour. Simple image patches (although in the YCbCr color space) are championed by [42], and a combination of simple features, such as corner positions and color, is used in [21], which is an area that warrants further research.

VII. DETECTION

The detection stage is where the signs are actually found. This is, in many ways, the most critical step and is often also the most complicated. The selection of detection method is a bit more constrained than the previous two stages since the method must work with the features from the previous stage. The decision is therefore often made the other way around: A desired detection method is chosen, and the feature extraction stage is designed to deliver what is necessary to perform the detection. As we know from the previous section, the most popular feature is the edges, and this reflects on the most popular choice in the detection method. Using Hough transforms to process the edges is one option, as done by [43] and [58]–[60]. In [60], a proprietary and undisclosed algorithm is used for the detection of rectangles, in addition to the Hough transform used for circles. That said, Hough transforms are computationally expensive and not suited for systems with real-time requirements. Because of that, the most popular methods are derivatives of the radial symmetry detector that was first proposed in [67] and first put to use for sign detection in [68]. The algorithm votes for the most likely sign centers in an image based on symmetric edges and is itself inspired by the Hough transform. The basic principle can be seen in Fig. 10. In a circle, all edge gradients intersect at the center. The algorithm finds gradients with a magnitude above a certain threshold. In the direction pointed out by the gradient, it casts a vote in a separate vote image. It looks for circles of a specific radius and thus votes only in the distance from the edge that is equivalent to the radius. The places with most votes are most likely to be the center of circles.



Fig. 11. Votes from a radial symmetry system superimposed on the original image. The brightest spot coincides with the center of the sign. This image is from a system developed in conjunction with this paper and is a radial symmetry voting algorithm extended to work for rectangles.

This algorithm was later extended to regular polygons by [35], and a faster implementation for sign detection use was proposed by [54]. It is also used in some form by [36], [38], [40], and [55]–[57]. An example of votes from a system that is extended to work for rectangular signs can be seen in Fig. 11. An alternate edge-based voting system is proposed by [61].

The HOG features can be used with a support vector machine (SVM), as in [19] and [65], or be compared by calculating a similarity coefficient, as in [17]. Another option with regard to HOG is to use a cascaded classifier trained with some type of boosting. This is done in [39] and [64]. Cascaded classifiers are traditionally used with Haar wavelets, and sign detection is no exception, as used in [37], [40], [51], and [66].

Finally, neural networks and genetic algorithms are represented in [42] and [49], respectively.

The detection stage reflects the philosophical difference that was also seen in the feature extraction stage: Either reliance on a simple theoretical model of sign shapes—at this stage, it is nearly always shapes that are searched for—or reliance on training data and, then, a more abstract detection method is preferred. Since it is extremely hard to compare systems tested across different data sets, it is not clear which methods perform the best; therefore, this is clearly an area that needs to be studied further. Both ways can be fast enough for real-time performance, and most of them could also work with signs of any shape. There are outliers using different methods, but there is no compelling argument that they should perform significantly better.

VIII. DISCUSSION AND FUTURE DIRECTIONS

In the previous sections, different methods and philosophies for each stage are presented. This section discusses the current state of the art and outlines ideas for future directions of research.

At the moment, the problem in TSR is the lack of use of standardized sign image databases. This makes comparisons between contributions very hard. To obtain meaningful advances

in the field, the development of such databases is crucial. Until now, research teams have only implemented a method that they believe has potential or perhaps tested a few solutions. Without a way to compare performance with other systems, it is not clear which approaches work best; therefore, every new team starts back at square one, implementing what *they* think might work best. Two efforts to remedy this situation deserve to be mentioned: The sign databases presented earlier and the segmentation evaluation in [7]. As mentioned earlier (see Section III-B), a few public sign databases have recently emerged but have not yet been widely used. In [7], the authors compare various segmentation methods on the same data set containing a total of 552 signs in 313 images. They also propose a way to evaluate the performance of segmentation methods. That paper provides a very good starting point for determining which segmentation method to use.

These two efforts notwithstanding, public databases covering signs from non-Vienna Convention regions are necessary. Databases that include video tracks of signs would also be very beneficial to the development of TSR systems since many detectors employ a tracking system for signs. This is, to some extent, included in the KUL Data set. In relation to the work on this present survey, we have assembled such a database for U.S. traffic signs, one that includes full video tracks of signs. It is our hope that the GTSRB database will also be extended to include video and full frames and that more U.S. databases will be created.

The absence of usage of public database may not explain in entirety why very few comparative studies of methods exist. Another reason is that TSR systems are long complex chains of various methods, where it is not always possible to swap individual modules. When it is not feasible to swap, for example, the detection method for something else, it is naturally hard to determine whether other solutions may be better. This is solved, if more papers divide their work more clearly into stages, ideally as fine grained as those used in this survey, plus a similar set of stages for classification. This is done with success in [7], as they test different segmentation methods while keeping the feature extraction, detection, and classification stages fixed.

Another problem is the need for work on TSR in regions not adhering to the Vienna Convention. The bulk of the existing work comes out of Europe, Australia, and Japan. Japan and Australia did not participate in the Vienna Convention, but they use similar signs, for example, to convey speed limits. Of the surveyed papers here, only two are concerned with U.S. traffic signs [40], [60], and even they only look at speed limit signs.

When looking at sign detection from a driver-in-the-loop perspective, it is also unfortunate that the bulk of research now focuses on speed limit signs. A wealth of papers cites driver assistance as their main application but carries on focusing on speed limit signs. Detection of speed limits is highly relevant for an autonomous vehicle, but as it turns out, humans are already very good at seeing speed limit signs themselves [9]. As such, recognition of signs other than speed limit is actually more interesting.

The final problem we wish to highlight in this section is the relation of signs to the surroundings. TSR has seen significant work, as is evident from this paper, but little work has been



Fig. 12. Example of sign relevancy challenges in a crop from our own collected data set. The signs have been manually highlighted, and while both signs would likely be detected, only the one to the right is relevant to the driver. The sign to the left belongs to another road, where the black and white cars come from.

done on ensuring that the detected signs are relevant for the ego car (with the notable exception of [58]). In many situations, it can occur that a detected sign is not connected to the road the car is on. An example from our own collected data can be seen in Fig. 12. In this case, two stop signs can be seen, but only the rightmost one pertains to the current road. Similar situations occur often on freeways, where some signs may only be relevant for exit lanes. Related to this problem is that, when the driver changes to a different road, most often, restrictions from earlier detected signs no longer apply. This should be detected and relayed to the system. It is very likely that research in other areas, such as lane detection can be of benefit here. Another idea with regard to the surroundings would be to link knowledge of weather and current lighting conditions to enhance the robustness of the detector, similar to what is done for detection of people in [69]. It is also possible that vehicle dynamics can be taken into account and used in the tracking of detected signs.

IX. CONCLUDING REMARKS

This paper has provided an overview of the state of sign detection. Instead of treating the entire TSR flow, focus has been solely on the detection of signs. In recent years, a lot of effort has gone into TSR, mainly from Europe, Japan, and Australia, and the developments have been described.

The detection process has been split into segmentation, feature extraction, and detection. Many segmentation approaches exist, mostly based on evaluating colors in various color spaces. For features, there is also a wealth of options. The choice is made in conjunction with the choice of detection method. By far, the most popular features are edges and gradients, but other options such as HOG and Haar wavelets have been investigated. The detection stage is dominated by the Hough transform and its derivatives, but for HOG and Haar wavelet features, SVMs, neural networks, and cascaded classifiers have also been used.

Arguably, the biggest issue with sign detection is currently the lack of use of public image databases to train and test systems. Currently, every new approach presented uses a new data set for testing, making comparisons between papers hard.

This gives the TSR effort a somewhat scattered look. Recently, a few databases have been made available, but they are still not widely used and cover only Vienna Convention-compliant i.e., signs. We have contributed with a new database, the LISA Data set, which contains U.S. traffic signs.

This issue leads to the main unanswered question in sign detection: Is a model-based shape detector superior to a learned approach, or *vice versa*? Systems using both approaches exist but are hard to compare since they all use different data sets.

Many contributions cite driver assistance systems as their main motivation for creating the system, but so far, only little effort has gone into the area of combining TSR systems with other aspects of driver assistance, and notably, none of the studies include knowledge about the driver's behavior to tailor the performance of the TSR system to the driver.

Other open issues include the lack of research into finding non-European style signs and the fact that detected signs are hard to relate to their surroundings.

ACKNOWLEDGMENT

The authors would like to thank our colleagues at the LISA-CVRR Laboratory, particularly S. Sivaraman, M. Van Ly, S. Martin, and E. Ohn-Bar for their comments.

REFERENCES

- [1] M. Trivedi, T. Gandhi, and J. McCall, "Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 1, pp. 108–120, Mar. 2007.
- [2] M. Trivedi and S. Cheng, "Holistic sensing and active displays for intelligent driver support systems," *Computer*, vol. 40, no. 5, pp. 60–68, May 2007.
- [3] C. Tran and M. M. Trivedi, "Vision for driver assistance: Looking at people in a vehicle," in *Guide to Visual Analysis of Humans: Looking at People*, T. B. Moeslund, L. Sigal, V. Krueger, and A. Hilton, Eds. New York: Springer-Verlag, 2011.
- [4] B. Morris and M. Trivedi, "Vehicle iconic surround observer: Visualization platform for intelligent driver support applications," in *Proc. IEEE IV Symp.*, 2010, pp. 168–173.
- [5] M.-Y. Fu and Y.-S. Huang, "A survey of traffic sign recognition," in *Proc. ICWAPR*, Jul. 2010, pp. 119–124.
- [6] H. Fleyeh and M. Dougherty, "Road and traffic sign detection and recognition," in *Proc. 10th EWGT Meet./16th Mini-EURO Conf.*, 2005, pp. 644–653.
- [7] H. Gomez-Moreno, S. Maldonado-Bascon, P. Gil-Jimenez, and S. Lafuente-Arroyo, "Goal evaluation of segmentation algorithms for traffic sign recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 4, pp. 917–930, Dec. 2010.
- [8] S. Houben, "A single target voting scheme for traffic sign detection," in *Proc. IEEE IV Symp.*, Jun. 2011, pp. 124–129.
- [9] D. Shinar, *Traffic Safety and Human Behaviour*. Bingley, U.K.: Emerald, 2007.
- [10] A. Doshi and M. Trivedi, "Attention estimation by simultaneous observation of viewer and view," in *Proc. IEEE Comput. Soc. Conf. CVPRW*, 2010, pp. 21–27.
- [11] E. Murphy-Chutorian, A. Doshi, and M. Trivedi, "Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation," in *Proc. IEEE ITSC*, 2007, pp. 709–714.
- [12] A. Doshi, S. Cheng, and M. Trivedi, "A novel active heads-up display for driver assistance," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 85–93, Feb. 2009.
- [13] A. De la Escalera, J. Armingol, and M. Mata, "Traffic sign recognition and analysis for intelligent vehicles," *Image Vis. Comput.*, vol. 21, no. 3, pp. 247–258, Mar. 2003.
- [14] *Convention on Road Signs and Signals of 1968*, United Nations Economic Commission for Europe, Geneva, Switzerland, 2006.
- [15] *California Manual on Uniform Traffic Control Devices for Streets and Highways*, State of California, Dept. Transp, Sacramento, CA, 2006.
- [16] A. Vázquez-Reina, S. Lafuente-Arroyo, P. Siegmann, S. Maldonado-Bascón, and F. Acevedo-Rodríguez, "Traffic sign shape classification based on correlation techniques," in *Proc. 5th WSEAS Int. Conf. Signal Process., Comput. Geometry Artif. Vis.*, 2005, pp. 149–154.
- [17] X. Gao, L. Podladchikova, D. Shaposhnikov, K. Hong, and N. Shevtsova, "Recognition of traffic signs based on their colour and shape features extracted using human vision models," *J. Vis. Commun. Image Represent.*, vol. 17, no. 4, pp. 675–685, 2006.
- [18] R. Timofte, K. Zimmermann, and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3D localisation," in *Proc. WACV*, 2009, pp. 1–8.
- [19] Y. Xie, L.-F. Liu, C.-H. Li, and Y.-Y. Qu, "Unifying visual saliency with HOG feature learning for traffic sign detection," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2009, pp. 24–29.
- [20] A. Ruta, F. Porikli, S. Watanabe, and Y. Li, "In-vehicle camera traffic sign detection and recognition," in *Machine Vision and Applications*. New York: Springer-Verlag, 2011, pp. 359–375. [Online]. Available: <http://dx.doi.org/10.1007/s00138-009-0231-x>
- [21] R. Kastner, T. Michalke, T. Burbach, J. Fritsch, and C. Goerick, "Attention-based traffic sign recognition with an array of weak classifiers," in *Proc. IEEE IV Symp.*, Jun. 2010, pp. 333–339.
- [22] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The german traffic sign recognition benchmark: A multi-class classification competition," in *Proc. IJCNN*, 2011, pp. 1453–1460. [Online]. Available: <http://benchmark.ini.rub.de/?section=gtsrb>
- [23] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition," *Neural Netw.*, vol. 32, pp. 323–332, Aug. 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0893608012000457>
- [24] R. Timofte, K. Zimmermann, and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3D localisation," in *Machine Vision and Applications*. New York: Springer-Verlag, Dec. 2011, pp. 1–15. [Online]. Available: <http://dx.doi.org/10.1007/s00138-011-0391-3>
- [25] F. Larsson and M. Felsberg, "Using Fourier descriptors and spatial models for traffic sign recognition," in *Proc. Image Anal.*, 2011, pp. 238–249.
- [26] C. Grigorescu and N. Petkov, "Distance sets for shape filters and shape recognition," *IEEE Trans. Image Process.*, vol. 12, no. 10, pp. 1274–1286, Oct. 2003.
- [27] R. Belaroussi, P. Foucher, J. Tarel, B. Soheilian, P. Charbonnier, and N. Paparoditis, "Road sign detection in images: A case study," in *Proc. ICPR*, Istanbul, Turkey, 2010, pp. 484–488.
- [28] R. Rajesh, K. Rajeev, K. Suchithra, V. Lekshesh, V. Gopakumar, and N. Ragesh, "Coherence vector of oriented gradients for traffic sign recognition using neural networks," in *Proc. IJCNN*, Aug. 5–31, 2011, pp. 907–910.
- [29] D. Ciresan, U. Meier, J. Masci, and J. Schmidhuber, "A committee of neural networks for traffic sign classification," in *Proc. IJCNN*, 2011, pp. 1918–1921.
- [30] F. Zaklouta, B. Stanculescu, and O. Hamdoun, "Traffic sign classification using K-d trees and Random Forests," in *Proc. IJCNN*, Aug. 5–31, 2011, pp. 2151–2155.
- [31] P. Sermanet and Y. LeCun, "Traffic sign recognition with multi-scale convolutional networks," in *Proc. IJCNN*, 2011, pp. 2809–2813.
- [32] F. Boi and L. Gagliardini, "A support vector machines network for traffic sign recognition," in *Proc. IJCNN*, 2011, pp. 2210–2216.
- [33] 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 Process.*, vol. 88, no. 12, pp. 2943–2955, Dec. 2008.
- [34] A. Ruta, Y. Li, and X. Liu, "Real-time traffic sign recognition from video by class-specific discriminative features," *Pattern Recognit.*, vol. 43, no. 1, pp. 416–430, Jan. 2010.
- [35] G. Loy and N. Barnes, "Fast shape-based road sign detection for a driver assistance system," in *Proc. IEEE/RSJ Int. Conf. IROS*, 2004, vol. 1, pp. 70–75.
- [36] B. Hoferlin and K. Zimmermann, "Towards reliable traffic sign recognition," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2009, pp. 324–329.
- [37] X. Baro, S. Escalera, J. Vitria, O. Pujol, and P. Radeva, "Traffic sign recognition using evolutionary adaboost detection and forest-ECOC classification," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 1, pp. 113–126, Mar. 2009.
- [38] Y. Gu, T. Yendo, M. Tehrani, T. Fujii, and M. Tanimoto, "Traffic sign detection in dual-focal active camera system," in *Proc. IEEE IV Symp.*, Jun. 2011, pp. 1054–1059.

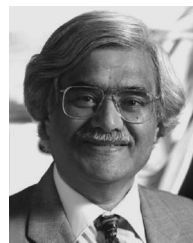
- [39] G. Overett and L. Petersson, "Large scale sign detection using HOG feature variants," in *Proc. IEEE IV Symp.*, Jun. 2011, pp. 326–331.
- [40] C. Keller, C. Sprunk, C. Bahlmann, J. Giebel, and G. Barattoff, "Real-time recognition of U.S. speed signs," in *Proc. IEEE IV Symp.*, Jun. 2008, pp. 518–523.
- [41] W.-J. Kuo and C.-C. Lin, "Two-stage road sign detection and recognition," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jul. 2007, pp. 1427–1430.
- [42] Y.-Y. Nguwi and A. Kouzani, "Detection and classification of road signs in natural environments," *Neural Comput. Appl.*, vol. 17, no. 3, pp. 265–289, Jun. 2008. [Online]. Available: <http://dx.doi.org/10.1007/s00521-007-0120-z>
- [43] F. Ren, J. Huang, R. Jiang, and R. Klette, "General traffic sign recognition by feature matching," in *Proc. 24th Int. Conf. IVCNZ*, Nov. 2009, pp. 409–414.
- [44] S. Xu, "Robust traffic sign shape recognition using geometric matching," *IET Intell. Transp. Syst.*, vol. 3, no. 1, pp. 10–18, Mar. 2009.
- [45] H.-H. Chiang, Y.-L. Chen, W.-Q. Wang, and T.-T. Lee, "Road speed sign recognition using edge-voting principle and learning vector quantization network," in *Proc. ICS*, Dec. 2010, pp. 246–251.
- [46] X. Qingsong, S. Juan, and L. Tiantian, "A detection and recognition method for prohibition traffic signs," in *Proc. Int. Conf. IASP*, Apr. 2010, pp. 583–586.
- [47] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, and F. López-Ferreras, "Road-sign detection and recognition based on support vector machines," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 2, pp. 264–278, Jun. 2007.
- [48] 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, no. 1, pp. 767–773, Jan. 2010. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1628324.1628558>
- [49] H. Liu, D. Liu, and J. Xin, "Real-time recognition of road traffic sign in motion image based on genetic algorithm," in *Proc. Int. Conf. Mach. Learn. Cybern.*, 2002, vol. 1, pp. 83–86.
- [50] X. Gao, K. Hong, P. Passmore, L. Podladchikova, and D. Shaposhnikov, "Colour vision model-based approach for segmentation of traffic signs," *J. Image Video Process.*, vol. 2008, pp. 6:1–6:7, Jan. 2008. [Online]. Available: <http://dx.doi.org/10.1155/2008/386705>
- [51] V. Prisacariu, R. Timofte, K. Zimmermann, I. Reid, and L. Van Gool, "Integrating object detection with 3D tracking towards a better driver assistance system," in *Proc. 20th ICPR*, Aug. 2010, pp. 3344–3347.
- [52] D. Deguchi, M. Shirasuna, K. Doman, I. Ide, and H. Murase, "Intelligent traffic sign detector: Adaptive learning based on online gathering of training samples," in *Proc. IEEE IV Symp.*, 2011, pp. 72–77.
- [53] P. Viola and M. Jones, "Robust real-time object detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2001.
- [54] N. Barnes and G. Loy, "Real-time regular polygonal sign detection," in *Field and Service Robotics*. New York: Springer-Verlag, 2006, pp. 55–66.
- [55] N. Barnes, A. Zelinsky, and L. Fletcher, "Real-time speed sign detection using the radial symmetry detector," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 2, pp. 322–332, Jun. 2008.
- [56] C. Nunn, A. Kummert, and S. Muller-Schneiders, "A two stage detection module for traffic signs," in *Proc. IEEE ICVES*, Sep. 2008, pp. 248–252.
- [57] M. Meuter, C. Nunn, S. M. Gormer, S. Muller-Schneiders, and A. Kummert, "A decision fusion and reasoning module for a traffic sign recognition system," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1126–1134, Dec. 2011.
- [58] M. A. Garcia-Garrido, M. Ocana, D. F. Llorca, M. A. Sotelo, E. Arroyo, and A. Llamazares, "Robust traffic signs detection by means of vision and V2I communications," in *Proc. 14th Int. IEEE ITSC*, Oct. 2011, pp. 1003–1008.
- [59] A. Gonzalez, M. Garrido, D. Llorca, M. Gavilan, J. Fernandez, P. Alcantarilla, I. Parra, F. Herranz, L. Bergasa, M. Sotelo, and P. Revenga de Toro, "Automatic traffic signs and panels inspection system using computer vision," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 485–499, Jun. 2011.
- [60] 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 *Proc. IEEE Intell. Veh. Symp.*, 2007, pp. 1122–1126.
- [61] R. Belaroussi and J.-P. Tarel, "Angle vertex and bisector geometric model for triangular road sign detection," in *Proc. WACV*, Dec. 2009, pp. 1–7.
- [62] A. Ruta, Y. Li, and X. Liu, "Towards real-time traffic sign recognition by class-specific discriminative features," in *Proc. 18th Brit. Mach. Vis. Conf.*, 2007, vol. 1, pp. 399–408.
- [63] B. Alefs, G. Eschemann, H. Ramoser, and C. Belezni, "Road sign detection from edge orientation histograms," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2007, pp. 993–998.
- [64] N. Pettersson, L. Petersson, and L. Andersson, "The histogram feature—A resource-efficient weak classifier," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2008, pp. 678–683.
- [65] I. Creusen, R. Wijnhoven, E. Herbschleb, and P. de With, "Color exploitation in hog-based traffic sign detection," in *Proc. 17th IEEE ICIP*, Sep. 2010, pp. 2669–2672.
- [66] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information," in *Proc. IEEE Intell. Veh. Symp.*, 2005, pp. 255–260.
- [67] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 959–973, Aug. 2003.
- [68] N. Barnes and A. Zelinsky, "Real-time radial symmetry for speed sign detection," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2004, pp. 566–571.
- [69] A. Doshi and M. Trivedi, "Satellite imagery based adaptive background models and shadow suppression," *Signal, Image Video Process.*, vol. 1, no. 2, pp. 119–132, Jun. 2007.



Andreas Møgelmoose received the B.Sc. degree in computer engineering on the topic of information processing systems and the Master's degree in vision, graphics, and interactive systems from Aalborg University, Aalborg, Denmark, in 2010 and 2012, respectively. He is currently working toward the Ph.D. degree with the Visual Analysis of People Laboratory, Aalborg University.

He has been a Visiting Scholar with the Computer Vision and Robotics Research Laboratory, University of California, San Diego. His research interests

are computer vision and machine learning.



Mohan Manubhai Trivedi received the B.E. (with honors) degree from Birla Institute of Technology and Science, Pilani, India, and the Ph.D. degree from Utah State University, Logan.

He is currently 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, University of California, San Diego (UCSD), La Jolla. He and his team are currently pursuing research on machine and human perception, machine learning, human-centered multimodal interfaces, intelligent transportation, driver assistance, and active safety systems. He serves as a consultant to industry and government agencies in the U.S. and abroad, including the National Academies, major auto manufacturers, and research initiatives in Asia and Europe.

Dr. Trivedi is a Fellow of the International Association for Pattern Recognition for contributions to vision systems for situational awareness and human-centered vehicle safety and a Fellow of The International Society for Optics and Photonics for distinguished contributions to the field of optical engineering. He was the Program Chair of the IEEE Intelligent Vehicles Symposium in 2006 and the General Chair of the IEEE IV Symposium in 2010. He is an Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS and the *Image and Vision Computing Journal*. He has been elected to the Board of members of the IEEE Intelligent Transportation Systems 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.



Thomas B. Moeslund received the M.Sc. and Ph.D. degrees in electrical engineering with a specialization in computer vision from Aalborg University, Aalborg, Denmark, in 1996 and 2004, respectively.

He is an Associate Professor and the Head of the Visual Analysis of People Laboratory, Aalborg University. He has published close to 100 peer-reviewed papers (citations: 2655. H-index: 16. Source: Harzing 24/1-2012). His research interests include all aspects of computer vision, with special focus on automatic analysis of people.

Mr. Moeslund serves as Associate Editor and Editorial Board member for four international journals and has cochaired four international workshops/tutorials. He has acted as program committee member/reviewer for a number of conferences and workshops. He has been involved in ten national and international research projects, both as coordinator, Work Package leader, and researcher. He received the Most Cited Paper Award in 2009, the Best IEEE Paper Award in 2010, and the Teacher of the Year Award in 2010.