

Traffic Sign Detection and Analysis: Recent Studies and Emerging Trends

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

Abstract—Traffic sign recognition (TSR) is a research field that has seen much activity in the recent decade. This paper introduces the problem and presents 4 recent papers on traffic sign detection and 4 recent papers on traffic sign classification. It attempts to extract recent trends in the field and touch upon unexplored areas, especially the lack of research into integrating TSR with a driver-in-the-loop system and some of the problems that presents. TSR is an exciting field with great promises for integration in driver assistance systems and that particular area deserves to be explored further.

I. INTRODUCTION

Traffic Sign Recognition (TSR) has seen much work in the past decade. With the emergence of increasingly complex Driver Assistance Systems (DAS), such as adaptive cruise control, including some sort of TSR for driver support has become a logical next step for inclusion in top-of-the-line cars. Some cars already come equipped with TSR for speed limit detection, but there are obviously many other signs that would be interesting to recognize from a DAS perspective.

The recent research in the field has been focused on the narrow vision-problem of detection, classification, and - to some extent - tracking of signs in images. For true integration in DAS, a TSR system should rather been looked upon as a driver-in-the-loop system where the driver is an integral part, as described in [1], [2], [3]. By also monitoring the driver, the system can tailor its output to specific situations. Furthermore, research indicates [4] that people are better at perceiving some signs than others, something that a TSR system could also benefit from taking into account to make sure that only relevant information is presented to the driver. There is not point in presenting a sign that the driver has already noticed.

TSR systems are traditionally split into a detection stage and a classification stage. The detection stage takes care of finding signs, while the classification stage figures out what a particular sign means. This paper describes each stage separately. It is possible to add a third stage that does tracking of the detected signs. The structure can be seen in fig. 1. The purpose of this paper is not to be a complete survey, but to highlight trends in the TSR research by using

A. Møgelmoose is a research scholar at the CVRR Lab, UCSD and PhD student at the VAP Lab, AAU, Denmark. andream@es.aau.dk

M. M. Trivedi is with the CVRR Lab at University of California, San Diego mtrivedi@ucsd.edu

T. B. Moeslund is with the VAP Lab at Aalborg University, Denmark tbm@create.aau.dk

some recent prominent papers as examples. The next section describes traffic signs along with some of the challenges and problem in detecting and recognizing them. After that is sections on how selected recent papers do the detection, classification, and tracking, respectively. That is followed up by a discussion of future directions in which the recent trends are examined and new or under-developed research areas are described.

II. ON TRAFFIC SIGNS

Traffic signs have the purpose of guiding people through the traffic in a safe manner. They are defined through laws, so the TSR task is quite well-defined. It is still, however, a complicated multi-class detection and classification problem, in some cases with extremely low intra-class variance.

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 [5]. 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 sub-signs in conjunction with one of the standard shapes. In addition to these, octagonal signs are used to signal a full stop, downwards pointing triangles yield and countries have different other types, e.g. to inform about city limits. Examples of these signs can be seen in fig. 2.

In the US, traffic signs are regulated by the Manual on Uniform Traffic Control Devices (MUTCD) [6]. 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, while the SHSM book has not been updated

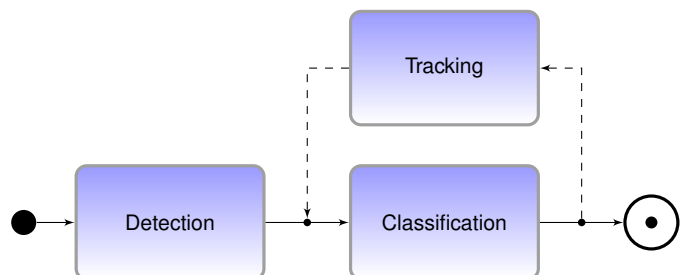


Fig. 1. The basic flow in most TSR systems.

since 2004, and thus it describes the MUTCD from 2003. An updated version of the SHSM should be on its way. The MUTCD contains a few hundred different signs, divided into 13 categories. US signs are white rectangles for regulatory signs, yellow diamonds for warnings, downwards pointing triangles for yield and octagons for full stop. Examples of American signs can be seen in fig. 3.

The Vienna Convention and the US MUTCD are the main standards. Most other countries use standards that are close to one of them, or a combination of the two. While signs seem to be well defined in many cases, the TSR task is made more difficult by a number of challenges. The signs may not be placed properly, so they are not perpendicular to the road, colors may be off due to wear or lighting conditions, they may be occluded by trees, poles, or other cars. Many signs, such as speed limit signs with different limits, are very similar to each other, making the classification task complicated.

III. DETECTION

As mentioned, the purpose of the detection stage is to find sign and pass them on to a classifier. It is common to treat detection and classification as two separate steps, but the interface between them is not standardized. Some classifiers rely on the detector to provide information on not only the center of the sign, but also its size, shape or overall sign type (e.g. regulatory sign vs. warning sign). Very often the attributes that determine the sign type - commonly shape and color - are also attributes the detector use, so this information is directly available.

Traditionally [12], [13], sign detectors have been classified as being either color-based or shape-based. Color-based detectors would find signs based on their distinctive background- or border-color, whereas shape-based detectors

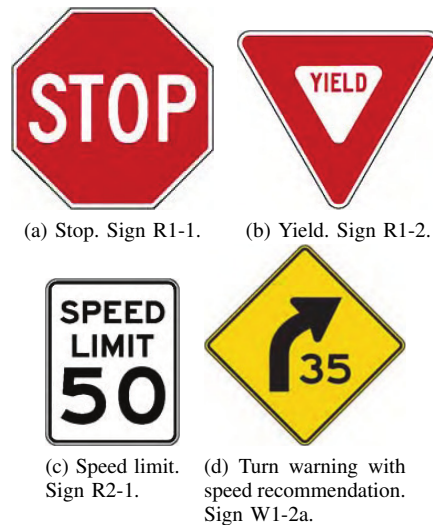


Fig. 3. Examples of signs from the US national MUTCD. Image source: [6]

would ignore color-information completely and find sign-shapes instead. This classification of detectors seem a bit outdated, since all color detectors also use shape information for further filtering. Champions of shape-based methods argue that color-detection is unreliable due to changes in lighting and sign wear. However, similar arguments can be put forth against shape-based detectors: Signs can be partly occluded or they may be rotated or otherwise distorted so their shapes look different, something not all shape based detectors can handle.

A better way to look at detectors is by splitting them into three blocks: Segmentation, feature extraction, and detection. Classification is not covered here, as that second part of the system is described in section IV. Almost all detection algorithms can be split into these blocks, making comparison across systems easy. Segmentation is usually color-based, but it may also be shape-based. It is the act of narrowing down the search to areas that are likely to contain signs. When that is done, features can be extracted from their areas. The choice of features is usually made in combination with the choice of the detector, since they work in unison to determine the actual signs.

In this paper, we have chosen to cover 4 recent leading papers [7], [9], [10], [11] that describe different methods of detecting signs. These papers, apart from being very recent, cover trends in the area well: Some use theoretical sign models, some use learned models, some are mainly color-based, some rely more on shapes, some have extensive focus on tracking. This means that they cover most directions in the field. An overview of the selected papers can be seen in table I. Each of the following subsections cover their methods used for the three blocks: Segmentation, feature extraction, and detection. For further analysis of traffic sign detection methods, see [14].

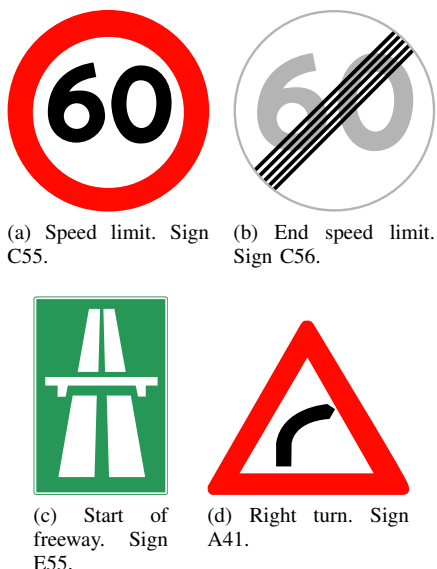


Fig. 2. Examples of European signs. These are Danish, but many countries use similar signs.

TABLE I
OVERVIEW OF DETECTION METHODS IN 4 RECENT PAPERS.

Paper	Year	Segmentation method	Features	Detection method
[7]	2010	HSI thresholding with addition for white signs ([8])	DtB (distance to bounding box)	Linear SVM
[9]	2011	Quad-tree color selection	Edges	Extended radial symmetry voting
[10]	2011	None	Various HOG-features	5 stage cascaded classifier trained with LogitBoost
[11]	2010	Biologically inspired attention model	Color, corner positions, height, eccentricity	Color, corner positions, height, eccentricity

A. Segmentation

[9] opts to use a color based segmentation. They propose a quad-tree attention operator. First step is a filtering that amplifies red and blue colors, the colors of the signs that the system is intended to work with. Then they compute a gradient magnitude map for each of the colors, and their corresponding integral images. Now, the image is evaluated for whether it contains a total color gradient over a certain threshold. If it does not, there is simply not enough colored edges in the sign to constitute any signs. If it does, the image is now split into four quarters, and the same check is done for each quarter. This process continues until a region goes below the threshold, or the minimum region size is reached. Adjacent regions that reach the minimum size while still containing enough gradients are clustered and constitute a sign candidate.

In [7], they follow the method described in their earlier paper, [15], and segment with a thresholding in the HSI (Hue, Saturation, Intensity) color space. It is argued that the HSI space is more robust to changes in lighting than the regular RGB (Red, Green, Blue) color space. They do, however add a method (originally pioneered by [8]), that finds achromatic colors and use this to find white signs. After the segmentation, image pixels that belong to the same color are grouped together.

[11] use a biologically inspired segmentation algorithm, which attempts to find areas in the image that are “interesting”. They compute an attention map based on various features, such as Difference of Gaussians (DoG), and Gabor filter kernels that mimics the brain of a mammal. This is done in the RGBY space, since that models how an eye works. These features are weighted and result in a map where high value areas are likely to contain signs.

In [10] they simply opt to not do any segmentation or preprocessing, but jump directly into feature extraction and detection.

For more on segmentation, see the great overview and comparison in [16].

B. Feature extraction

The features that must be extracted must be chosen in close connection with the detection method. In [9], they test both an edge based detector and a cascade using Haar-like feature [17], but end up using the edge based one. Thus, their features are simply the image gradients.

The detector in [7] relies on Distance to Bounding box (DtB) features. It is a measure of distances from the edges of

an object to its rectangular bounding box. A rectangular sign will have zero distance to its bounding box, while an upwards pointing triangle will have zero distance to the bottom of its bounding box, but increasing distances when approaching to the upper corners of the bounding box.

To obtain features in [11], they run a color thresholding and then calculate a number of geometric features, such as corner positions, size and eccentricity.

In [10], two different types of Histogram of Oriented Gradient (HOG) features are used. HOG features are, as the name suggests, histograms detailing the orientation of the gradients in an area. Thus, all horizontal lines are binned together, as are vertical lines, etc.

C. Detection

The detection block is where the features for each sign candidate are evaluated and it is determined whether they describe a sign or not. The detection can either be done by matching a theoretical model with the feature (such as deciding whether the candidate looks like a circle), or by matching the features with a learned model of how signs should look in these particular features.

[9], [11] use a theoretical model. In [9], a center-voting scheme based on circles’ edges, first presented in [18], is used to find sign candidates. [11] use a template for where corners should be located.

[7], [10] instead use learned classifiers. [7] use a Support Vector Machine (SVM) classifier on the DtB features and [10] use a similar cascaded classifier, trained with LogitBoost.

IV. CLASSIFICATION

Classification is where the meaning of the detected signs are determined. It is a classical computer vision task. Recently, the competition “The German Traffic Sign Recognition Benchmark” (GTSRB) [23] has put renewed focus on the classification. It is a competition with the objective of classifying a number of German (and thus Vienna Convention compliant) signs in no less than 43 classes. The number of classes alone makes this a challenging task. The competition attracted many competitors and spawned four papers [19], [20], [21], [22] from the best competitors. These papers can be said to represent the state-of-the-art in sign classification. An overview can be seen in table II. They achieve very good classification rates for the GTSRB dataset.

TABLE II
OVERVIEW OF CLASSIFICATION METHODS IN THE 4 PAPERS FROM THE GTSRB CONTEST.

Paper	Year	Features	Classification method	Classification rate
[19]	2011	Hue histograms and HOG	Network of SVM classifiers	96.89%
[20]	2011	48x48 pixel color normalized image patches	Convolutional neural network	98.98%
[21]	2011	32x32 pixel image patches in the YUV color space	Convolutional neural network	98.97%
[22]	2011	HOG-features	K-d trees and random forests	97.2%

Unlike the detection task, where some systems employ a theoretical model instead of a learned one, all competitors used a learned classifier. [19] use a network of SVM classifiers. It runs a preprocessing to normalize and enhance colors and calculate the features used: A set of hue histograms and a set of HOG-features. [20] - the winner of the competition - use a Convolutional Neural Network (CNN) and does not extract specific features, but use full 48x48 pixel color normalized image patches. A CNN is inspired by the primary visual cortex [24] and described further in [25], [26]. [21] also use a convolutional network on full image patches, this time resized to 32x32 pixels and converted to the YUV color space. [22] use K-d trees (similar to [27]) with the Best Bin First algorithm described in [28] and random forests on HOG-features.

V. TRACKING

Tracking is the act of following a sign through several frames. Tracking is not used by any of the papers mentioned in the classification section above, since they were simple passed an image of a sign and could leave any tracking to the detector. Detectors, however, can benefit vastly from incorporating a tracking algorithm. Not only can it be used to discard false positives by discarding signs that only appear in a single frame - usually the result of noise - they can also use it to only present new signs to the classifier, enhancing the speed of the system. Furthermore, a sophisticated tracking system can make sure that signs that are temporarily occluded are not reported as new signs when they show up again.

Of the selected papers, only one employ tracking: [9]. It has a sophisticated tracking system based on the changes in appearance of the sign. When detecting a sign, it is assumed to be undistorted. Then a number of random deformations of that particular sign is generated. These distorted views are used to train the tracker on the fly. The motion is learned by fitting these to the sign in following frames using regression. The system is described further in [29], [30].

VI. DISCUSSION AND FUTURE DIRECTIONS

TSR is an area that has seen a lot of contributions recently, and it is an area that is well researched. The main shortcoming is that for detection, no standardized dataset is used, so comparison among papers is hard. Only a few public datasets exist that are suitable for detection: The Swedish Traffic Signs Dataset [31], the KUL Belgium Traffic Sign Classification Benchmark [32], and most recently the LISA Dataset [14]. None of them are widely used yet. The lack of common datasets was recently

remedied for the classification stage, where the GTSRB dataset is a good contribution which is already used in a few papers. For training purposes, synthetic images has also recently been explored in [33], but is was deemed unsatisfactory, thus underlining the need for these datasets.

The trends seem to be towards more thoroughly tested and compared systems. This effort is spearheaded by the GTSRB, but something similar is needed for detection. It also seems that the trend goes toward learned systems rather than pre-programmed heuristics. Earlier, the common thing has been to create full systems covering both detection and classification, but with the GTSRB, systems has been more modularized and it has become common to create systems that only do classification, something that will make it easier to mix and match approaches to arrive at a system that is fit for a specific application.

However, when looking at TSR in a bigger perspective, much remains to be done. Good detection and classification systems exist, but little work on how to apply TSR in actual systems exist. As mentioned in the introduction, many TSR systems cite driver assistance as their motivation, but simply recognizing signs does not help the driver. In order for TSR to be really applied to driver-in-the-loop systems, it is crucial to take him into account. One option is to look at driver attention: Why present the driver with signs that he has already seen? That will only contribute to information overload. It may also be necessary to pay special attention to signs that drivers are known to simply glance over, as presented in [4].

For a driver-in-the-loop system tracking becomes even more crucial than it already is. As of now, it is mostly used to increase robustness, or not at all. When a driver is present, it is important not to present the same sign to him twice, again to prevent information overload. This means that when a sign is temporarily occluded, it should be handled by tracking so it is not discovered as a new sign when it shows up again. There is also the issue of how to present recognized signs to the driver. In general, the area of really including the driver in TSR systems are virtually unexplored.

VII. CONCLUDING REMARKS

This paper has presented 4 significant recent papers in the area of sign detection and 4 in the area of classification. TSR systems have seen much activity recently, but progress is hampered by the fact that comparison across papers is

hard when no standardized dataset for detection exists. Still, very good systems show up, and especially the classification seems to fare very well. This is helped by the new image database, the GTSRB.

Still, much research remains to be done in the area of applying TSR to DAS. Proper integration of the two is a very promising and exiting task that is in need of much more attention. While many systems perform well in the area when viewed strictly as an object detection or classification task, not much work has been done in applying such systems to driver assistance.

VIII. ACKNOWLEDGMENT

The authors would like to thank our colleagues in the LISA-CVRR lab.

REFERENCES

- [1] M. Trivedi, T. Gandhi, and J. McCall, "Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8, no. 1, pp. 108–120, 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, 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., 2011.
- [4] D. Shinar, *Traffic safety and human behaviour*. Emerald Group Publishing, 2007.
- [5] United Nations Economic Commission for Europe, "Convention on Road Signs And Signals, of 1968," 2006.
- [6] State of California, Department of Transportation, "California Manual on Uniform Traffic Control Devices for Streets and Highways."
- [7] S. Lafuente-Arroyo, S. Salcedo-Sanz, S. Maldonado-Bascón, J. A. Portilla-Figueras, and R. J. López-Sastre, "A decision support system for the automatic management of keep-clear signs based on support vector machines and geographic information systems," *Expert Syst. Appl.*, vol. 37, pp. 767–773, January 2010. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1628324.1628558>
- [8] H. Liu, D. Liu, and J. Xin, "Real-time recognition of road traffic sign in motion image based on genetic algorithm," in *Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on*, vol. 1. IEEE, 2002, pp. 83–86.
- [9] A. Ruta, F. Porikli, S. Watanabe, and Y. Li, "In-vehicle camera traffic sign detection and recognition," *Machine Vision and Applications*, vol. 22, pp. 359–375, 2011. [Online]. Available: <http://dx.doi.org/10.1007/s00138-009-0231-x>
- [10] G. Overett and L. Petersson, "Large scale sign detection using HOG feature variants," in *Intelligent Vehicles Symposium (IV), 2011 IEEE*, June 2011, pp. 326–331.
- [11] R. Kastner, T. Michalke, T. Burbach, J. Fritsch, and C. Goerick, "Attention-based traffic sign recognition with an array of weak classifiers," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*, June 2010, pp. 333–339.
- [12] M.-Y. Fu and Y.-S. Huang, "A survey of traffic sign recognition," in *Wavelet Analysis and Pattern Recognition (ICWAPR), 2010 International Conference on*, July 2010, pp. 119–124.
- [13] H. Fleyeh and M. Dougherty, "Road and traffic sign detection and recognition," in *10th EWGT Meeting and 16th Mini-EURO Conference*, 2005, pp. 644–653.
- [14] A. Møgelmoose, M. M. Trivedi, and T. B. Moeslund, "Vision based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey," *IEEE Intelligent Transportation Systems Transactions and Magazine*, vol. Special Issue on MLFTSR, Dec 2012.
- [15] 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," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8, no. 2, pp. 264–278, 2007.
- [16] H. Gomez-Moreno, S. Maldonado-Bascon, P. Gil-Jimenez, and S. Lafuente-Arroyo, "Goal Evaluation of Segmentation Algorithms for Traffic Sign Recognition," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 11, no. 4, pp. 917–930, Dec. 2010.
- [17] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2001.
- [18] G. Loy and N. Barnes, "Fast shape-based road sign detection for a driver assistance system," in *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 1. IEEE, 2004, pp. 70–75.
- [19] F. Boi and L. Gagliardini, "A Support Vector Machines network for traffic sign recognition," in *Neural Networks (IJCNN), The 2011 International Joint Conference on*. IEEE, 2011, pp. 2210–2216.
- [20] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, "A committee of neural networks for traffic sign classification," in *Neural Networks (IJCNN), The 2011 International Joint Conference on*. IEEE, 2011, pp. 1918–1921.
- [21] P. Sermanet and Y. LeCun, "Traffic sign recognition with multi-scale Convolutional Networks," in *Neural Networks (IJCNN), The 2011 International Joint Conference on*. IEEE, 2011, pp. 2809–2813.
- [22] F. Zaklouta, B. Stanculescu, and O. Hamdoun, "Traffic sign classification using K-d trees and Random Forests," in *Neural Networks (IJCNN), The 2011 International Joint Conference on*, 31 2011-Aug. 5 2011, pp. 2151–2155.
- [23] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German Traffic Sign Recognition Benchmark: A multi-class classification competition," in *Neural Networks (IJCNN), The 2011 International Joint Conference on*. IEEE, 2011, pp. 1453–1460. [Online]. Available: <http://benchmark.ini.rub.de/?section=gtsrb>
- [24] D. Hubel and T. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *The Journal of physiology*, vol. 148, no. 3, pp. 574–591, 1959.
- [25] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [26] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?" in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 2146–2153.
- [27] W.-J. Kuo and C.-C. Lin, "Two-Stage Road Sign Detection and Recognition," in *Multimedia and Expo, 2007 IEEE International Conference on*, July 2007, pp. 1427–1430.
- [28] J. Beis and D. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. IEEE, 1997, pp. 1000–1006.
- [29] E. Bayro-Corrochano and J. Ortégón-Aguilar, "Lie algebra approach for tracking and 3D motion estimation using monocular vision," *Image and Vision Computing*, vol. 25, no. 6, pp. 907–921, 2007.
- [30] O. Tuzel, F. Porikli, and P. Meer, "Learning on lie groups for invariant detection and tracking," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.
- [31] F. Larsson and M. Felsberg, "Using fourier descriptors and spatial models for traffic sign recognition," *Image Analysis*, pp. 238–249, 2011.
- [32] R. Timofte, V. Prisacariu, L. Van Gool, and I. Reid, "Combining Traffic Sign Detection with 3D Tracking Towards Better Driver Assistance," *Emerging topics in computer vision and its applications*, 2011.
- [33] A. Møgelmoose, M. M. Trivedi, and T. B. Moeslund, "Learning to Detect Traffic Signs: Comparative Evaluation of the Roles of Real-world and Synthetic Datasets," in *21st International Conference on Pattern Recognition*, Nov 2012.