Large-scale classification of traffic signs under real-world conditions

Lykele Hazelhoff[a,b], Ivo Creusen[a,b], Dennis van de Wouw[a,b] and Peter H.N. de With[a,b]

[a]CycloMedia Technology B.V., Achterweg 38, 4181 AE, Waardenburg, The Netherlands;
[b]Eindhoven University of Technology, Den Dolech 2, 5600 MB, Eindhoven,The Netherlands

ABSTRACT

Traffic sign inventories are important to governmental agencies as they facilitate evaluation of traffic sign locations and are beneficial for road and sign maintenance. These inventories can be created (semi-)automatically based on street-level panoramic images. In these images, object detection is employed to detect the signs in each image, followed by a classification stage to retrieve the specific sign type. Classification of traffic signs is a complicated matter, since sign types are very similar with only minor differences within the sign, a high number of different signs is involved and multiple distortions occur, including variations in capturing conditions, occlusions, viewpoints and sign deformations. Therefore, we propose a method for robust classification of traffic signs, based on the Bag of Words approach for generic object classification. We extend the approach with a flexible, modular codebook to model the specific features of each sign type independently, in order to emphasize at the inter-sign differences instead of the parts common for all sign types. Additionally, this allows us to model and label the present false detections. Furthermore, analysis of the classification output provides the unreliable results. This classification system has been extensively tested for three different sign classes, covering 60 different sign types in total. These three data sets contain the sign detection results on street-level panoramic images, extracted from a country-wide database. The introduction of the modular codebook shows a significant improvement for all three sets, where the system is able to classify about 98% of the reliable results correctly.

Keywords: Classification, Object Detection, Bag of Words, Traffic Sign Recognition, Inventory Systems

1. INTRODUCTION

Nowadays, several companies record street-level panoramic images, which give an accurate and recent overview of the road infrastructure. Within The Netherlands, CycloMedia captures these images annually, recording every 5 meters on each public road, resulting in a detailed and recent overview of the local situation. The resulting databases can be exploited for inventories of street furniture, including traffic signs. Such inventories are of interest to governmental institutes tasked with the maintenance of the roads and street furniture, in order to maintain a high road safety, which is directly influenced by the accurate placement of traffic signs. Moreover, the visibility of traffic signs may be lowered due to aging, vandalism, accidents or vegetation coverage. Therefore, up-to-date inventories of traffic signs are of high importance for these instances, especially when the signs with lowered visibility are included and marked correspondingly. Such inventories can be performed by hand, tracking all roads, but efficiency can be improved by exploiting the mentioned panoramic images, and additionally by automatization based on computer-vision techniques. We have set up such a system, which applies object detection for identifying the traffic signs present in the images, followed by a classification stage to obtain the specific sign type for each detection. Afterwards, for the classified traffic signs, the position coordinates are computed by exploiting the position data during capturing.

Automatic detection and classification of traffics signs is a complicated problem for several reasons. First, variations in capturing conditions exist (including lighting conditions, sharpness, occlusions). Second, viewpoint changes occur (such as scaling, rotation, skew), which are inherent due to the capturing with a moving car. Third, traffic signs may be less visible due to aging, vandalism, accidents or vegetation coverage, while especially these signs are interesting for indicating sign maintenance. Some examples are shown in Fig. 1. Fourth, many similar traffic signs exist, having the same shape and color, but with only minor differences in the inner sign part. Moreover, also custom versions of official signs are sometimes placed, which have the same function, but a slightly different appearance. Examples of these cases are shown in Fig. 2. For these real-world conditions, we have found

(a) (b) (c) (d) (e) (f) (g) (h)

Figure 1. Examples of traffic signs with lowered visibility. Although these situations occur infrequently, these cases are most interesting for road safety.

(a) (b) (c) (d) (e) (f) (g) (h)

Figure 2. Examples of traffic signs which show highly similarities with other sign types (a)-(f), together with examples of a standard version (g) and custom variant (h) of a warning sign.

that the performance of state-of-the-art algorithms for detection and classification are insufficient to perform a completely automatic inventory of traffic signs, similar as Frome *et al.*[1] found for face detection. Therefore, a semi-automatic approach is chosen to ensure a highly accurate inventory, which is required for professional applications, as reported by Creusen *et al.*[2] Although our sign detector,[3] which exploits very generic sign properties based on an implicit shape model, obtains most sign classes (e.g. red triangular, blue circular) with a high accuracy, discrimination between the often very similar sign types (e.g. speed limit signs with different text) is rather difficult. For a semi-automatic system, we aim at obtaining a high accuracy, but not necessarily for all samples, as we have the possibility of manual intervention. This prompts for algorithms that improve the inventory efficiency, in particular by marking unreliable results such that manual interaction can be applied specifically to these samples, whereas reliable results may be accepted immediately. This paper presents such an algorithm for robust classification of traffic signs, aiming at the reduction of the required amount of manual interaction. Besides marking of possibly unreliable results, we additionally require a generic system, that is applicable to several sign classes without major modifications.

The described classification system is based on the generic Bag of Words (BoW)[4] approach, a popular technique for object classification. In this technique, characteristics of small pixel regions are mathematically described, using a so-called descriptor function. In literature, numerous descriptor functions are known, including SIFT,[5] SURF,[6] DAISY[7] and HoG,[8] where dense sampling outperforms interest point operators, as shown by Nowak *et al.*[9] Since the amount of extracted descriptors is extremely large, a relatively small codebook is formed, aiming at containing the most common descriptors in the training examples (called visual words). A common approach is to apply K-means clustering, which places clusters near frequently occurring descriptors. When the visual dictionary is obtained, the image is described by constructing a histogram of visual word occurrences, by matching each extracted descriptor to the closest visual word. Based on these histograms, a supervised classification scheme is trained, for which numerous techniques are available,[10] including Neural Networks, Adaboost and Support Vector Machines. After training, the classification process for an unseen sample involves matching each extracted descriptor to the closest visual words, construction of the word histogram and classification based on this histogram.

This paper describes a classification system based on BoW, where three modifications to the BoW approach are introduced to adjust the system to the specific situation of traffic signs. The first modification deals with the high similarities between the different sign types within the same sign class, since e.g. the sign borders are equal for all types. Within the standard BoW approach, the visual dictionary contains visual words located near descriptor features that occur frequently over all the training samples. However, these words may not be the most discriminative.[11] This will especially be the case when these words occur in all sign types, such as e.g. words representing the sign borders, as they are frequently occurring, but not discriminative. Therefore, a separate dictionary is generated for each individual sign type, which also enables handling of unbalanced data and reduction of the computational costs during dictionary generation. Afterwards, all these individual codebooks are combined to a single, large visual dictionary, referred to as the *modular codebook*.

The second modification deals with the false detections given by the traffic sign detectors. The chosen approach with a modular codebook containing individual parts for each sign type allows for the addition of a specific dictionary part representing these false detections, such that patterns occurring frequently in false detections will be mapped to this part of the modular dictionary. The modular codebook is used for classifier training and classification, where we employ an All-Versus-One classification structure, equipped with linear Support Vector Machines.

The third modification involves marking of samples belonging to none or multiple classes as unreliable. Since the system is designed for a semi-automatic procedure, this allows us to direct the manual interaction to both those samples and the samples classified as false detections, while accepting signs types for which a high classification score is achieved. This system is extensively tested for three different Dutch sign classes: red triangular warning sings, red circular restriction signs and blue circular direction signs, which all contain a large number of slightly differing sign types. We will show that our approach has resulted in a generic algorithm, applicable to multiple sign classes and with an increased performance w.r.t. the standard BoW approach. Moreover, our algorithm can handle the false detections present in our system and outputs classifications that are possibly unreliable, and thereby contributes to a more efficient traffic sign inventory system.

The remainder of the paper is organized as follows. Section 2 describes the algorithm in detail, including both the training and the evaluation stages. The performed experiments and results are described in Section 3, followed by the conclusions and future work in Section 4.

2. ALGORITHMIC DESCRIPTION

2.1 System overview

The system described in this paper is based on the BoW approach, where we apply several modifications to adjust the system to the specific case of traffic signs. This system consists of two parts: the training stage and the evaluation stage, which are briefly described below.

- *Training stage:* In this stage, the system learns the recognition of a certain sign type. For this, features that describe local pixel neighborhoods are extracted from each training sample, based on which the visual dictionary is created. We do this for each sign type independently, and append all obtained dictionaries to a single dictionary. Then, the occurrences of the obtained visual words in this dictionary are represented in a word histogram, one for each training sample. Based on this, a One-vs-All classification structure is trained to discriminate between the different sign types.
- *Evaluation stage:* In the evaluation stage, the system labels an unseen sample. For this, the same feature extraction is applied as during the training phase, and each feature is mapped to the closest visual word contained in the visual dictionary. The occurrences of all words are counted in the word histogram, where after the trained classification structure flags whether each target class is present. Results are marked as unreliable in case not a single, or multiple sign types are recognized (so everything except a single sign type).

2.2 Training Stage

In the training stage, SIFT features are extracted from all training samples, using a dense grid. Since we are interested in the sign contents, each training sample is manually segmented, and only descriptors that do not fall outside the sign contour are taken into account, as shown in Fig. 3. Our system exploits grayscale SIFT features, which are chosen for several reasons. First of all, the dissimilar parts of the different sign types are mostly in black and white as the signs are designed to be also visible to color blind people. Secondly, grayscale SIFT features have shown a high robustness to varying recording conditions. After the SIFT features are extracted for all training samples, we apply K-means clustering to obtain the K features that are most prominent. Since our data is highly unbalanced, while the difference between certain sign types is rather small, we apply these steps to each sign type independently, using $K = 50$. So, for each sign type a visual dictionary is generated independently, and afterwards, all these dictionaries are concatenated to the modular codebook, which thus contains all the individual dictionary parts. Next to insensitivity to unbalanced data, this approach has additional advantages, including (1) a high flexibility, since additional sign types can easily be added and removed from the modular codebook, and (2) lowered computation requirements, as less features are subject to clustering at the same time. Furthermore, this approach also allows modeling of the occurring false detections by adding these as an additional class, where the approach enables using a larger dictionary size for them by selecting e.g. $K = 250$. This enables assigning false detections a specific label, while preventing mapping them to a random sign type. The modular codebook structure is portrayed by Figure 4.

The obtained modular codebook is applied in the same way as the regular codebook in the standard BoW approach. For every training sample, the extracted SIFT features are mapped to the closest code word, and the matches are counted in the word histogram, which is $L2$ normalized. Based on the histograms of all training samples, we train a classification structure, where we apply the commonly used One-versus-All (OvA) approach. We employ linear Support Vector Machines for classification.

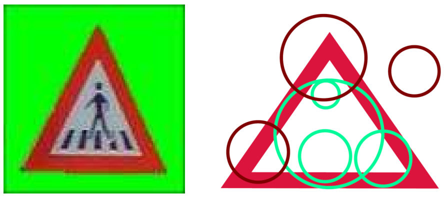

Figure 3. Segmented training image and visualization of descriptors used for training (green) and descriptors that are discarded during training (maroon).

Figure 4. Illustration of the modular codebook: a concatenation of multiple smaller codebooks representing the different sign classes together with a larger codebook representing false detections.

2.3 Evaluation Stage

The evaluation stage aims at classifying an unseen input sample into one of the target sign types. For this, at first descriptors are extracted from the input image, again using a dense grid. Afterwards, each descriptor is matched to the closest visual word present in the modular codebook, and the matches are counted in the word histogram. Afterwards, the word histograms are $L2$ normalized and used for classification, where the classification approach trained above is employed for assigning a label to the unseen sample. Since we apply an One-Versus-All approach, each classifier outputs whether a certain sign type is present, and ideally, only a single classifier should give a positive result. When none or multiple classifiers attempt to assign a label to a sample, we mark the result as unreliable, which is beneficial in our semi-automatic approach, as this enables us to filter and check these samples in particular.

3. EXPERIMENTS AND RESULTS

3.1 Data set description

The described algorithm is validated for three different sign classes, i.e. red triangular warning signs, red circular restriction signs and blue circular direction signs. For each class, a data set is created, containing the raw output of our traffic sign detector[3] for that specific sign class. The detector outputs nearly all target signs in a size

Figure 5. Examples of traffic sign cut-outs present in our data set.

range of 24×24 pixels up to 400×400 pixels, corresponding to a sign-to-camera distance of about 19 to 1 meter. Since a new image is captured every 5 m, each physical sign is recorded multiple times at different sizes. As signs captured from a large distance have a rather low resolution, their inner part is not expected to be discriminative, and therefore, we ignore samples smaller than 40×40 pixels and focus on the samples with higher resolution. Due to the nature of traffic signs, the resulting data sets are highly unbalanced, as not all signs occur frequently. Therefore, the sets are constructed such that as much as possible different sign types are included. This resuls in data sets with sufficient cardinality for 27, 20 and 13 sign types (out of the 39, 29 and 15 officially defined types), respectively. As the detector is tuned to find as much signs as possible, also a number of false detections is obtained, which are included in the data sets.

We should note that these street-level panoramic images are captured under different weather conditions in different seasons and with different cameras, resulting in data sets with challenging variations. Since the images are captured from a driving car, motion blur may also be present. Figure 5 displays examples from our data sets.

3.2 Experimental setup

The described algorithm is applied to the three above-described data sets, together with an implementation of the standard BoW approach, using a codebook obtained by K-means clustering and OvA classification. Based on this, we conduct two experiments. First, we compare our approach to the standard BoW approach without the presence of false detections. Second, we investigate the classification performance of our approach when false detections are present. These experiments are discussed in the sections below. During these experiments, the evaluation is carried out by applying 10-fold cross validation, as this approaches the systems performance when all samples were subject to training.[12] This involves splitting of the data sets into 10 equally sized, non-overlapping partitions, where the system is alternately tested using one selected partition, after training with the 9 other partitions. Since our data set contains multiple samples of the same physical sign, captured from different viewpoints, these occurrences are forced into the same partition to prevent testing with the training set. These folds are kept constant during all experiments.

The experiments result in confusion matrices, which show the occurrence of the classification output for all sign types, together with the results labeled as unreliable (placed in the last column). The performance of our classification system is numerically assessed based on these matrices. For this assessment, we calculate the percentage of classifications that is correct, i.e. the *correct classification rate*, and we do this with and without counting the unreliable samples. In this way, we are able to both evaluate the performance of the classification system for a fully automated operation mode, where no samples are checked, and for a semi-automatic environment, where the samples labeled as unreliable, are subject to manual evaluation.

3.3 Comparison of our approach with the standard BoW approach

Table 1 shows the results of the comparison of our approach with the standard BoW approach, where in both systems the size of the visual dictionaries are chosen equally large. As follows from the table, the usage of the modular codebook significantly improves the classification results, and our approach outperforms the standard

	Red triangular signs		Red circular signs		Blue circular signs	
	Or. BoW	Ours	Or. BoW	Ours	Or. BoW	Ours
# samples	$16,010$	$16,010$	$7,565$	$7,565$	$15,935$	$15,935$
# correctly classified samples	$14,361$	$14,925$	$6,927$	$7,204$	$15,432$	$15,464$
# falsely classified samples	879	318	409	102	344	256
# unreliable samples	770	767	229	259	159	215
Correct classification rate	90.2%	93.6%	91.6%	95.2%	96.8%	97.0%
Correct class. rate wo unreliable	94.2%	97.9%	94.4%	98.6%	97.8%	98.4%

Table 1. Performance overview for the three sign classes for the original BoW and our approach.

BoW approach for all three sign types. When analyzing the incorrect classifications, we have found that they are often caused by large perspective deformations and by sign damages, e.g. through vandalism or accidents. When unknown samples are not counted, our modifications allow a correct classification rate about 98% for all reliably classified samples, which is an acceptable score in our inventory system. The reader should note the difference in improvement between the red triangular and red circular signs on one hand, and blue circular signs on the other hand. We suspect that this is caused by the nature of the red-sign classes, as they only differ slightly in the inner part and have large common regions. This feature does not appear in the blue circular signs, which are more distinctive w.r.t. each other.

For completion, we have included the confusion matrices for the red circular signs for both systems in Table 3 and Table 4, respectively.

3.4 Classification with false detections

The described classification algorithm is intended to work on the output of the traffic sign detector, which is tuned to identify almost all present signs, thereby also outputting a number of false detections. To cope with such detections, we have treated them as an additional sign type by including a large visual dictionary, representing the false detections, to the modular codebook. Here, we investigate the influence of the false detections on the classification procedure. Table 2 displays the classification results for the proposed approach. The table contains also the results for the case that the background is not modeled, using the classifiers from Sect. 3.3. Ignoring the false detections during the classifier training stage results in poor performance, where 63% of the occurring false detections are classified as a sign type (and the rest as unreliable). In contrast, including false detections to the classification procedure enables appropriate labeling of about 92.4% of them, where only about 2.5% of them are assigned a sign type label. The classification score for the real signs is slightly lowered, where about 1.3% of the signs are labeled as background. However, the classification accuracy is still approximately 97%, such that the output of the classifier is still considered to be reliable.

For clarity, we have included the confusion matrices for the blue circular signs with and without incorporating background to the training procedure in Table 5 and Table 6, respectively.

4. CONCLUSIONS AND FUTURE WORK

This paper has presented an algorithm for robust classification of traffic signs, aiming at improving the efficiency of a semi-automatic inventory system. However, classification of traffic signs in real world circumstances is a difficult problem, since the sign appearances are often disturbed and different sign types only differ in minor details. Therefore, we desire a system that explicitly models the output reliability, such that possible errors can be corrected manually, while reliable output is accepted immediately. For this purpose, we have designed a classification algorithm, based on the popular Bag of Words approach for generic object classification. We have extended the BoW approach with a modular codebook, containing a fixed number of codewords per class, ensuring that all sign-specific features are covered, instead of the overall commonly occurring features. This modular codebook also allows for the addition of a specific type, modeling the false detections, which prevents false detections being assigned the closest sign type in many cases. Furthermore, we label unreliable results, which enables specific manual intervention to augment reliability.

	Red triangular signs		Red circular signs		Blue circular signs	
	No FD class	With FD class	No FD class	With FD class	No FD class	With FD class
# total samples	17,344	17,344	10,296	10,296	19,319	19,319
# total sign samples	16,010	16,010	7,565	7,565	15,935	15,935
# total FD samples	1,406	1,406	2,731	2,731	3,384	3,384
# correctly classified samples	14,925	16,067	7,204	9,617	15,464	18,470
# falsely classified samples	1,309	849	1,859	270	2,446	513
# unreliable samples	1,110	428	1,233	409	1,409	336
# signs classified as FD	0	126	0	134	0	265
# FD classified as sign	792	43	1,757	44	2,190	104
All samples:						
Correct classification rate	86.1%	92.6%	70.0%	93.4%	80.0%	95.6%
Correct class. rate wo unreliable	93.1%	97.4%	79.5%	97.3%	86.3%	97.3%
Signs only (no FD):						
Correct classification rate	93.6%	93.1%	95.2%	93.2%	97.0%	96.1%
Correct class. rate wo unreliable	97.9%	97.5%	98.6%	96.9%	98.4%	97.4%

Table 2. Performance overview for the three sign classes for the case that false detections (FD) are ignored during classifier training, compared to the case that the background is modeled as if it would it be a sign type.

The algorithm is extensively evaluated for three different Dutch sign types: red triangular warning signs, red circular restriction signs and blue circular direction signs. Compared to the standard BoW approach, the system shows a significant improvement in the number of correctly classified samples. When only taking the reliably classified samples into account (which is of importance for our application), the described system achieves a correct classification rate of about 98%, and thereby outperforms the standard BoW system. A further benefit is that our system successfully recognizes most false detections, so that the remaining manual interaction is guided to specific suspicious samples in our semi-automatic traffic sign inventory system.

For the future, we work on the construction of larger data sets for evaluation featuring larger nubmer of classes and samples. Next to this, it is interesting to investigate the classification consistency per real traffic sign, as each sign is recorded 3 to 4 times, which possibly may increase the classification accuracy.

REFERENCES

[1] Frome, A., Cheung, G., Abdulkader, A., Zennaro, M., Wu, B., Bissacco, A., Adam, H., Neven, H., and Vincent, L., "Large-scale privacy protection in google street view," in [*Proc. IEEE International Conference on Computer Vision (ICCV)*], 2373–2380 (October 2009).

[2] Creusen, I., Hazelhoff, L., and With, P. D., "A semi-automatic traffic sign detection, classification and positioning system," in [*Proceedings of SPIE Volume 8305-25*], (2012).

[3] Creusen, I., Wijnhoven, R., Herbschleb, E., and de With, P., "Color exploitation in hog-based traffic sign detection," in [*ICIP*], (2010).

[4] Csurka, G., Dance, C. R., Fan, L., Willamowski, J., and Bray, C., "Visual categorization with bags of keypoints," in [*Proc. European Conference on Computer Vision (ECCV)*], (May 2004).

[5] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," *Int. Journal of Computer Vision (IJCV)* **60** (January 2004).

[6] Bay, H., Tuytelaars, T., and van Gool, L., "Surf: Speeded up robust features," in [*LNCS , vol. 7 ch. 395112006, 2006*], 404–417 (2006).

[7] Tola, E., Lepetit, V., and Fua, P., "A fast local descriptor for dense matching," in [*IEEE CVPR*], (2008).

[8] Dalal, N. and Triggs, B., "Histogram of oriented gradients for human detection," in [*IEEE CVPR, Vol. 1*], 886–893 (June 2005).

[9] Nowak, E., Jurie, F., and Triggs, B., "Sampling strategies for bag-of-features image classification," in [*Proc. European Conference on Computer Vision (ECCV)*], 490–503, Springer (2006).

[10] Bishop, C., [*Pattern Recognition and Machine Learning*], Springer (2006).

[11] Jurie, F. and Triggs, B., "Creating efficient codebooks for visual recognition," in [*ICCV*], 604–610 (2005).

[12] Kohavi, R., "A study of cross-validation and bootstrap for accuracy estimation and model selection," in [*IJCAI*], (1995).

	30	50	60	70										3,1m	4,8t	5,4t				P	UR
30	837	10	7	2	0	0	0	0	0	3	0	2	0	4	0	4	0	0	0	0	24
50	10	524	11	0	0	0	1	0	0	2	0	2	0	2	1	3	0	1	0	0	22
60	10	14	832	0	1	0	1	0	0	5	0	2	0	2	1	3	0	0	0	0	22
70	3	0	0	268	0	0	0	0	0	1	2	1	0	2	0	1	0	1	0	0	6
	1	0	0	0	879	0	1	0	1	1	0	22	0	2	1	1	3	0	0	0	9
	0	1	0	0	0	91	4	0	0	1	0	0	0	0	0	0	0	0	0	0	5
	1	2	4	0	0	0	651	0	0	2	0	0	0	1	0	2	0	0	0	2	14
	0	0	0	0	0	0	1	48	0	0	0	1	0	0	0	0	0	0	0	0	2
	1	0	0	0	0	0	0	0	378	5	2	9	0	0	0	0	0	0	0	0	13
	2	4	3	0	0	1	0	0	2	669	0	12	0	9	2	6	0	0	0	0	19
	0	1	1	0	0	0	0	0	0	3	145	2	0	3	0	1	0	0	0	0	11
	1	2	1	0	6	0	0	0	3	11	2	336	0	7	6	2	1	0	0	0	17
	2	2	0	0	1	0	0	0	1	3	1	1	171	0	0	1	2	0	0	0	4
3,1m	0	0	1	1	1	0	1	0	0	3	1	10	0	146	4	5	1	0	0	0	15
4,8t	2	1	1	0	0	0	0	0	0	4	0	9	0	4	222	0	0	0	0	0	16
5,4t	8	2	6	1	0	0	0	0	1	3	0	4	0	1	0	86	0	0	0	0	14
	0	3	0	0	1	0	0	0	1	3	0	3	1	0	0	1	229	0	0	0	7
	0	2	0	0	0	0	0	0	1	2	0	2	0	0	0	1	0	187	0	0	7
	0	0	1	0	0	0	0	0	0	3	0	1	0	0	0	1	0	0	125	0	1
P	0	0	4	0	0	0	0	0	1	5	0	2	0	0	0	0	0	0	0	103	1

Table 3. Confusion Matrix of the red circular signs and the standard BoW approach.

	30	50	60	70										3,1m	4,8t	5,4t				P	UR
30	865	4	2	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	20
50	6	537	9	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	24
60	2	1	855	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	33
70	1	0	0	280	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
	0	0	1	0	900	0	0	0	0	0	4	1	2	3	0	0	0	0	0	0	10
	0	0	1	0	0	95	1	0	0	0	0	0	0	0	0	0	0	0	0	0	5
	0	0	0	0	0	0	659	0	0	0	0	0	0	0	0	1	0	0	0	0	19
	0	0	0	0	0	0	1	49	0	0	0	0	0	0	0	0	0	0	0	0	2
	0	0	0	0	0	0	0	0	400	0	0	0	0	0	0	0	0	0	0	0	8
	1	0	1	0	0	0	0	0	0	703	0	2	0	1	1	0	0	0	0	0	20
	1	0	0	0	0	0	0	0	0	0	151	1	0	1	0	1	0	0	0	0	12
	0	0	1	0	2	0	0	0	0	3	2	358	0	1	1	0	0	0	0	0	27
	0	0	0	0	0	0	0	0	0	1	0	0	180	0	0	0	0	0	0	0	8
3,1m	1	0	0	0	3	0	0	0	0	0	1	2	0	160	2	2	0	0	0	0	18
4,8t	2	0	0	0	0	0	0	0	0	2	1	1	0	1	241	0	0	0	0	0	11
5,4t	5	0	5	0	0	0	0	0	0	0	0	0	0	3	0	85	0	0	0	0	28
	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	243	0	0	0	5
	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	199	0	0	1
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	131	0	1
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	113	3

Table 4. Confusion Matrix of the red circular signs and our approach with the modular codebook.

	FD														UR
FD	0	41	559	17	3	9	0	0	203	24	794	26	485	29	1194
	0	854	3	0	0	0	0	0	0	0	1	0	1	1	6
	0	0	5089	0	0	0	0	1	5	0	16	0	12	2	38
	0	0	4	442	0	0	0	0	1	0	1	0	2	0	6
	0	0	2	0	112	0	0	0	0	0	2	0	1	0	1
	0	0	2	0	0	116	0	0	1	0	1	0	0	0	3
	0	0	1	0	0	0	85	0	0	0	3	0	0	0	2
	0	0	4	0	0	0	0	100	0	1	1	0	1	0	5
	0	0	7	0	0	0	0	0	1500	0	12	1	6	1	20
	0	0	1	0	0	0	0	0	0	43	2	0	6	0	4
	0	0	9	0	0	0	0	0	4	0	3762	3	36	0	63
	0	1	0	0	0	0	0	0	0	0	16	139	6	1	9
	0	0	5	0	0	0	0	0	2	0	36	2	3062	6	47
	0	0	0	0	0	0	0	0	1	0	4	0	18	160	11

Table 5. Confusion matrix for the blue circular signs where the false detections (FD) are not considered during classifier training. UR denotes unreliable.

	FD														UR
FD	3161	1	21	1	0	0	0	0	16	1	42	2	18	2	119
	17	840	0	0	0	0	0	0	0	0	2	0	0	0	7
	70	1	5054	0	0	0	0	1	3	0	6	0	3	0	25
	9	0	3	437	0	0	0	0	3	0	0	0	0	0	4
	4	0	0	0	110	0	0	0	0	0	2	0	0	0	2
	1	0	1	0	0	116	0	0	0	0	0	0	0	0	5
	2	0	1	0	0	0	81	0	2	0	0	0	0	0	5
	3	0	4	0	0	0	1	98	0	0	1	0	0	0	5
	35	0	0	0	0	0	0	0	1481	0	1	0	2	0	28
	7	0	1	0	0	0	0	0	0	43	1	0	0	0	4
	68	0	4	0	0	0	0	0	0	0	3732	3	22	0	48
	6	0	0	0	0	0	0	0	0	0	10	136	6	2	12
	37	0	1	0	0	0	0	0	2	0	26	0	3022	7	65
	6	0	0	0	0	0	0	0	0	0	2	0	20	159	7

Table 6. Confusion matrix for the blue circular signs where the false detections (FD) are modeled as a sign type, and are included in the modular codebook. UR denotes unreliable.