Semi-supervised Training Set Adaption to Unknown Countries for Traffic Sign Classifiers

Matthias Hillebrand[1], Christian Wöhler[2], Ulrich Kreßel[1], and Franz Kummert[3]

[1] Daimler AG, Group Research and Advanced Engineering, 89081 Ulm, Germany
[2] Image Analysis Group, TU Dortmund, 44221 Dortmund, Germany
[3] Applied Informatics Group, Bielefeld University, 33615 Bielefeld, Germany
{matthias.hillebrand,ulrich.kressel}@daimler.com,
christian.woehler@tu-dortmund.de,
franz@techfak.uni-bielefeld.de

Abstract. Traffic signs in Western European countries share many similarities but also can vary in colour, size, and depicted symbols. Statistical pattern classification methods are used for the automatic recognition of traffic signs in state-of-the-art driver assistance systems. Training a classifier separately for each country requires a huge amount of training data labelled by human annotators. In order to reduce these efforts, a self-learning approach extends the recognition capability of an initial German classifier to other European countries. After the most informative samples have been selected by the confidence band method from a given pool of unlabelled traffic signs, the classifier assigns labels to them. Furthermore, the performance of the self-learning classifier is improved by incorporating synthetically generated samples into the self-learning process. The achieved classification rates are comparable to those of classifiers trained with fully labelled samples.

Keywords: Pattern recognition, self-training, sample selection, confidence bands.

1 Introduction

Many stationary or mobile systems depend on sensory perception of their environment. Intensity-based classifiers are commonly utilised for processing visual sensor information. This contribution considers the automatic recognition of traffic signs by a driver assistance system.

The traffic signs in Western European countries reveal only varieties regarding colour, font, font size and depicted symbols. In a classical supervised learning-based approach, a classifier has to be trained for each country separately. But such an approach is inefficient because of the high labelling costs for human annotators, while unlabelled data can often be acquired done with justifiable expenditure, for example by driving camera-equipped cars and applying an automatic detection algorithm.

A straightforward approach to this problem begins with an initial training set for one country and extends it with the most informative samples from other

F. Schwenker and E. Trentin (Eds.): PSL 2011, LNAI 7081, pp. 120–127, 2012.

countries. An automatic class assignment is desirable to further reduce the labelling costs. This approach is known in literature as semi-supervised learning.

We propose an iterative training process where the most informative samples from a given pool of unlabelled traffic signs are selected by the confidence band method so that a label can be automatically assigned by the classifier itself. Additional knowledge of noise distributions in rotation, camera angle, etc., obtained from the initial training set is provided to the selection algorithm as a set of synthetically generated samples.

The approach evaluation starts with a fully labelled training set of German traffic signs and adapts the system to traffic signs from Austria and Switzerland without any intervention by a human expert. The classifier distinguishes between 12 different classes of traffic signs including speed limits, no-overtaking signs, and the corresponding ending signs. In the following sections, the proposed method is elaborated in detail.

2 Related Work and Applied Methods

2.1 Prerequisites

First, a detection algorithm generates hypotheses from grey-value intensity camera images by applying the Hough method for circles. The second step consists of a normalisation including a resizing of all images to 17×17 pixels and the adjustment of lighting conditions. The intensity values are then used as a feature vector of dimension 289 reduced to 25 by a Principal Component Analysis preserving 81% of the image information.

Since the evaluation only considers traffic sign patterns for the classifier training, the set of hypotheses is divided into a non-sign (garbage) and a sign set in the next step. For this purpose we utilise a second-order polynomial classifier [8] trained on the initial German traffic sign set. The classification leads to a false positive rate of about 5%, which means that each 20th pattern classified as a sign is actually a garbage sample. Similar classification rates for the sign and garbage division on the Austrian and Swiss data sets result from using the classifier trained on the German set.

2.2 Traffic Sign Recognition

Recognition of traffic signs is a mature but still a contemporary field of research. The survey by Fu and Huang [3] provides an introduction and a brief overview of the broad field of existing approaches. Today, traffic sign recognition systems are available as special equipment for some cars of renowned manufacturers. None of these systems have self-learning capabilities.

2.3 Classification

Not all classifiers are equally well-suited for our classification task. The self-learning process nearly always associates a certain fraction of samples with the wrong classes. Different classifier architectures show considerable differences in their sensitivity towards these mislabelled samples. For instance, standard Support Vector Machines (SVM) are very sensitive against outliers and mislabelled training samples [11] while classifiers with a neural network like architecture relying on continuous input, output, and weight values can be trained with partially mislabelled data and are nonetheless capable of producing good classification results [10]. This study utilises a polynomial classifier [8] with a fully quadratic structure of the feature vectors. The decision in favour of this classifier was due to several reasons: First, the training is fast. Second, the classifier is robust to partially mislabelled data. Finally, re-training can be performed easily by mixing old and new moment matrices.

2.4 Self-learning

In the iterative process of training self-learning classifiers, new samples are selected from the large set of available unlabelled samples. The selected samples are then classified or rejected by the classifier, and if not rejected, are added to the training set along with their classifier-predicted labels. At the end of each iteration, the classifier is re-trained with the extended training set and the training procedure is repeated. The surveys by Zhu and Goldberg [12] and Chapelle et al. [1] give a comprehensive overview of the field of semi-supervised learning techniques, also the one described above, commonly referred to as self-training according to Zhu. To the authors' knowledge, self-training methods have not been applied yet to the field of traffic sign recognition.

2.5 Sample Selection

The crucial step in the iterative training process of a semi-supervised or self-learning classifier is the selection of the most informative samples which are to be added to the training set during each iteration in order to re-train the classifier. In self-learning processes without a human teacher, the classifier must be capable of determining labels for the selected unlabelled samples.

An overview to selection methods is given in the survey by Settles [9]. Common approaches are uncertainty sampling [6] and the confidence value estimation from Conditional Random Fields [2]. More recently, the concept of confidence bands was applied to self-learning classifiers for handwritten digits and traffic signs in [4].

2.6 Confidence Bands

Confidence bands are curves enclosing a model function being estimated by a regression analysis. The bands represent the areas where the true model is expected to reside with a certain probability, commonly 95%. The extent of the bands in different areas of the data space gives an idea of how well the estimated function fits the data.

While Martos et al. [7] describe an analytical approach for computing prediction bands in a camera calibration application, Hillebrand et al. [4] adapt the algorithm to compute the closely related confidence bands for application in the context of polynomial classification or regression: A confidence band value is computed for each sample. During each iteration a reference confidence band value (i.e., the average value) is determined from all labelled samples. Based on the minimum difference of their band values to the reference value, a maximum of n samples (e.g., $n = 100$) is selected, classified and added to the training set. The selection of samples with confidence band values close to the average value avoids redundancies (low values) on the one hand and the selection of samples from feature sub-spaces with too much model uncertainty (high values) on the other.

2.7 Virtual Training Samples

Real images of traffic signs depict a wide spectrum of variations, e.g., different sizes, rotations and camera angles, translations due to inaccurate detection results, soiling, partial occlusions, or different lighting and weather conditions. Lighting conditions are normalised by a pre-processing algorithm. The other most common variations (size, rotation, camera angle, translation) are represented by virtual traffic sign samples which are generated from one ideal depiction of each traffic sign by a method described by Hoessler et al. [5], but the less frequent variations are unrepresented. In principle, an infinite amount of such virtual traffic sign samples are available.

3 Experimental Evaluation

3.1 Experimental Setup

The performance evaluation applies the adaptive self-learning classifier to different learning scenarios. As a basis, the classifier is trained with a fully labelled set of German traffic sign samples. This data set consists of 12 classes (see Fig. 1) containing 500 samples per class (6000 samples in total). A training set of the same composition and nearly the same size is available from Austria (5987 samples in total because some classes do not comprise the full number of samples). Furthermore, a smaller training set of only 4428 samples which are not equally distributed over all classes is available from Switzerland (four classes are underrepresented, especially the class *speed limit 70 km/h* with only 11 samples). We refer to the classifier trained with the German training set as *German reference classifier*. The same naming convention applies to the *Austrian* and the *Swiss reference classifier*.

Finally, one set of virtual training samples for each country, again with 12 classes and 6000 samples in total, is created and referred to as the virtual classifiers, e.g., as the *German virtual classifier*.

The classifier performances are compared by computing correct classification rates and false classification rates on independent test sets. Like the training

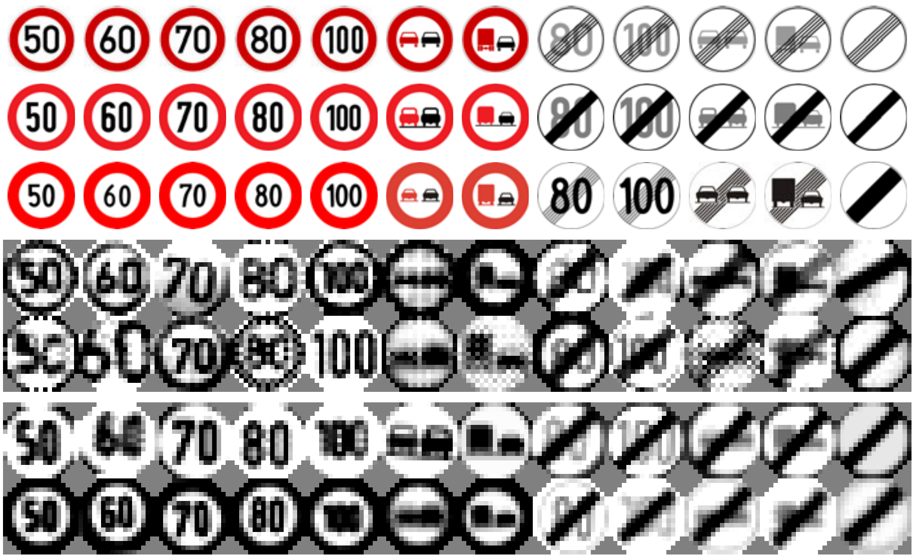

Fig. 1. Traffic sign images. Rows 1–3: Ideal depictions of 12 German, Austrian, and Swiss traffic signs. Rows 4–5: Two German real training samples of each class. Rows 6–7: Two Austrian virtual training samples of each class.

sets, the test sets consist of samples from the same 12 classes. Each class is represented by 250 samples (3000 samples in total) which have been recorded independently with different cameras. The same pre-processing routines have been applied to all samples (real training, virtual training, test).

The selection of German training samples depicted in Fig. 1 shows that the set is not noisy-free and contains a considerable amount of garbage samples and bad quality samples, e.g., with inaccurate cutouts. The proportion of these garbage and bad quality samples is between 5% and 10%. The same applies to the Austrian and Swiss sets. Classifiers often obtain correct classification rates above 90%, so the large number of bad quality samples in the test sets would not allow meaningful comparisons between different classifiers. For that reason, all samples not classifiable by a human expert have been removed from the test sets before the 3000 samples per country were chosen randomly.

For comparison, the performance of the German reference classifier on the German, Austrian, and Swiss test sets as well as the performance of the Austrian reference classifier on the Austrian test set and the Swiss classifier on the Swiss test set were determined. These measurements were repeated with the corresponding virtual classifiers.

As expected, the results presented in Table 1 indicate that the reference classifiers outperform the virtual classifiers in Germany and Austria because some variations described in Section 2.7 are not represented by the virtual training samples. In Switzerland, the virtual classifier performs much better than the

Table 1. Correct classification rates of German, Austrian, and Swiss reference and virtual classifiers on the different test sets. All values are given in percent.

classifier	German		Austrian		Swiss	
	ref.	virt.	ref.	virt.	ref.	virt.
German test set	97.4	89.9				
Austrian test set	88.1	77.8	96.3	87.7		
Swiss test set	85.3	73.1			89.2	95.4

reference classifier due to missing training samples for some classes and the high number of bad quality samples.

3.2 Self-learning with Real Traffic Signs

The German reference classifier performs 88.1% on the Austrian and 85.3% on the Swiss test set; the performances of the Austrian and Swiss virtual classifiers amount to 87.7% and 95.4%, respectively. The objective is for the generated self-learning classifiers to have a higher recognition performance.

The training process starts with an initial set containing all fully labelled German training samples. Then the training continues iteratively as described by adding Austrian samples which have been labelled by the classifier itself. Our classifier outperforms the German reference classifier (88.1%) after some iterations and once all training samples have been added, a performance of 93.1% is reached.

The difference to the Austrian reference classifier (using the same training samples) can be explained by the presence of garbage samples in the training set (about 5% as described in Section 2.1) that cannot be assigned a “correct” class label. Furthermore, a certain fraction of the added samples (about 15%) are mislabelled by the classifier and influence the training process negatively, especially when appearing in the early stages of the process. Making use of both of these samples in self-training will inevitably twist the feature distribution of each class.

A completely different behaviour can be observed when training the Swiss self-learning classifier: the performance (81.5%) is even lower than the performance of the German reference classifier (85.3%). This is likely due to the missing training samples and the bad quality of the existing ones, which results in a high fraction of mislabelled samples (around 31% of the added samples).

Further experiments vary the initial training set sizes by adding 10% (600 samples) of the labelled Austrian and Swiss samples, respectively. As a result, the self-learning classifiers obtain much better performances due to a lower rate of mislabelled samples (around 15% and 29%, respectively): the Austrian self-learning classifier improves to 95.3% and the Swiss classifier to 84.4%. Of course, the performance increases come at the price of being dependent on a human labelling expert to some extent again.

Adding another supplemental 10% labelled Austrian samples (now 1200 samples in total) to the initial training set results in a decrease of the rate of mislabelled samples by another 0.5% but has nearly no improving effect on the classification rates. With the additional 10% labelled samples, the Swiss classifier improves marginally to 85.2% and the rate of mislabelled samples decreases to 27%, but this performance is still worse than that of the German reference classifier.

3.3 Self-learning with Virtual Traffic Signs

When disposing of a theoretically endless supply of virtual training samples, it appears reasonable to extend the German standard classifier with a huge amount of Austrian respectively Swiss virtual training samples. After adding 6000 virtual training samples each, classifier performances of 95.2% (Austrian) and 94.9% (Swiss) are reached. Clearly, the classification rate of the Austrian classifier is nearly equal to the one of the self-learning classifier trained before, while the performance of the Swiss classifier is nearly equal to that of the Swiss virtual classifier.

Finally, the self-learning classifiers are combined with the recently trained virtual classifiers. The initial training sets are constructed from all 6000 German fully labelled samples and all 6000 Austrian and Swiss virtual samples, respectively, also fully labelled. Then we start the self-learning process by adding all Austrian respectively Swiss real training samples iteratively. These classifiers perform with correct classification rates of 95.7% and 95.2%, respectively, which are the best performances except for the Austrian reference classifier and the Swiss virtual classifier. The remarkable point here is that these high performances were reached without any intervention by a human labelling expert.

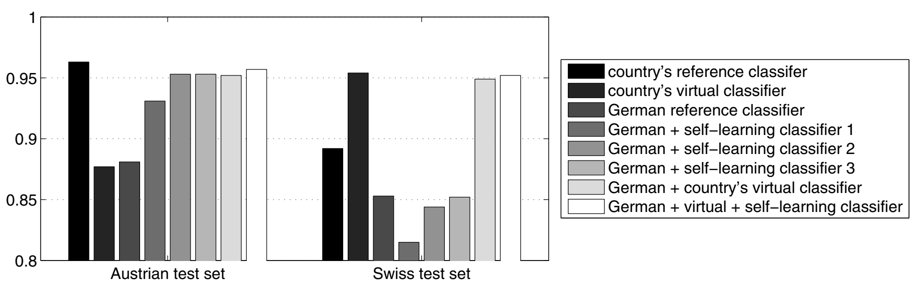

Fig. 2. Classification rates. Comparison of the correct classification rates of the classifiers described in Sections 3.2 and 3.3 on the Austrian and the Swiss training set.

4 Summary and Conclusions

The proposed self-learning classification system is capable of adapting itself to changed appearances of known traffic signs from other countries. This reduces the labelling efforts by human expert annotators and therefore the overall costs significantly.

The Swiss classifier achieves its best performance when trained with Swiss virtual samples only. This behaviour is due to the fact that the Swiss training set contains a huge amount of bad quality samples. A comparison with other countries reveals that a classifier exclusively trained based on virtual samples is always capable of classifying traffic sign images to a certain degree correctly.

The Austrian classifier achieves the best performance when trained with virtual samples first and subsequently self-trained with real samples. Since a Swiss classifier trained in this way performs nearly as good as its virtual classifier, this strategy is a suitable compromise.

References

1. Chapelle, O., Schölkopf, B., Zien, A. (eds.): Semi-Supervised Learning. Adaptive Computation and Machine Learning. The MIT Press (2006)
2. Culotta, A., McCallum, A.: Confidence Estimation for Information Extraction. In: Proc. of Human Language Technology Conference and North American Chapter of the Association for Computational Linguistics (HLT-NAACL), pp. 109–112 (2004)
3. Fu, M.Y., Huang, Y.S.: A survey of traffic sign recognition. In: Proc. of the International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), pp. 119–124 (2010)
4. Hillebrand, M., Wöhler, C., Krüger, L., Kreßel, U., Kummert, F.: Self-learning with confidence bands. In: Proc. of the 20th Workshop Computational Intelligence, pp. 302–313 (2010)
5. Hoessler, H., Wöhler, C., Lindner, F., Kreßel, U.: Classifier training based on synthetically generated samples. In: Proc. of the 5th International Conference on Computer Vision Systems (ICCV) (2007)
6. Jeon, J.H., Liu, Y.: Semi-supervised Learning for Automatic Prosodic Event Detection Using Co-training Algorithm. In: Proc. of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, pp. 540–548 (2009)
7. Martos, A., Krüger, L., Wöhler, C.: Towards Real Time Camera Self Calibration: Significance and Active Selection. In: Proc. of the 4th Int. Symp. on 3D Data Processing, Visualization and Transmission (3DPVT) (2010)
8. Schürmann, J.: Pattern Classification: A Unified View of Statistical and Neural Approaches. John Wiley & Sons (1996)
9. Settles, B.: Active Learning Literature Survey. Computer Sciences Technical Report 1648. University of Wisconsin–Madison (2010)
10. Wöhler, C.: Autonomous in situ training of classification modules in real-time vision systems and its application to pedestrian recognition. Pattern Recognition Letters 23(11), 1263–1270 (2002)
11. Xu, L., Crammer, K., Schuurmans, D.: Robust Support Vector Machine Training via Convex Outlier Ablation. In: Proc. of the 21st National Conference on Artificial Intelligence (AAAI), pp. 536–542 (2006)
12. Zhu, X., Goldberg, A.B.: Introduction to Semi-Supervised Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers (2009)