Selection of an Optimal Set of Discriminative and Robust Local Features with Application to Traffic Sign Recognition

Benjamin Höferlin
Universität Stuttgart, Germany
benjamin.hoeferlin@vis.uni-stuttgart.de

Gunther Heidemann
Universität Stuttgart, Germany
ais@vis.uni-stuttgart.de

ABSTRACT

Today, discriminative local features are widely used in different fields of computer vision. Due to their strengths, discriminative local features were recently applied to the problem of traffic sign recognition (*TSR*). First of all, we discuss how discriminative local features are applied to TSR and which problems arise in this specific domain. Since TSR has to cope with highly structured and symmetrical objects, which are often captured at low resolution, only a small number of features can be matched correctly. To alleviate these issues, we provide an approach for the selection of discriminative and robust features to increase the matching performance by speed, recall, and precision. Contrary to recent techniques that solely rely on density estimation in feature space to select highly discriminative features, we additionally address the question of features' retrievability and positional stability under scale changes as well as their reliability to viewpoint variations. Finally, we combine the proposed methods to obtain a small set of robust features that have excellent matching properties.

Keywords: Discriminative Local Features, Traffic Sign Recognition, SIFT.

1 INTRODUCTION

Over the past years, local features have become the preferred method in different fields of computer vision. Today they are applied in many tasks like panoramic imaging [BL07] or object recognition [Low04]. In this work we especially focus on scale-invariant and discriminative local features like the popular *SIFT* (scale-invariant feature transform) [Low04] or *SURF* (Speeded up robust features) [BTG06]. Such local features are commonly calculated in two stages. First, an interest point detector is used to find salient image regions at their characteristic scale. Then, a robust descriptor is extracted for each region.

Their ability to represent an image by the means of local patch descriptors, might be the reason for their success. Using local features, two images can be compared very fast, since only some salient regions are considered instead of the whole images. Besides this, local features like SIFT are more robust to various image transformations than global methods are. On the one hand, this is due to the extraction of image patches around salient interest points using a scale-invariant detector. On the other hand, the feature description is often covariant to a variety of image changes, too, including rotation or lighting changes. In contrast to global methods, local features can inherently cope with partial occlusions in the image. As result to their depicted strengths, discriminative local features have found their way into the field of traffic sign recognition (*TSR*), too. In literature, TSR is often split into the problems of traffic sign detection (TSD) and traffic sign classification (TSC). These applications are highly relevant to many recent advanced driver assist systems (ADAS), since the information annotated to the streets is primarily visually encoded in traffic signs. Due to the fact, that drivers tend to trust in the information provided by ADAS [BT05], TSR has to be reliable. This demand grows with the emergence of active ADAS-technologies like brake-by-wire. The demand for reliability leads to the application of local features for their illustrated vantages. But local feature approaches also involve weaknesses, like slow matching performance on large databases or mismatches due to features with poor saliency. In this work, we address these shortcomings and propose a method able to cope with them.

In detail, our contribution is the introduction of a novel method for the selection of a small but highly discriminative set of local features for TSR. We do not solely consider their discriminative power regarding a single image, but also their saliency regarding the traffic sign domain. In addition to their high saliency, the selected features have to be positionally stable and robust under large scale changes. Also the features' stability under viewpoint changes is evaluated in order to reduce the influence of features violating the local planarity assumption. Finally, a greedy algorithm is introduced to select a small and optimal set of local features according to their properties. Although, the feature set can be of arbitrary size, we go for a small amount to achieve fast feature matching.

The remainder of this paper is structured as follows: First, we briefly introduce how local features are applied to TSR and which specific problems arise within the traffic sign domain. In section 3, we then discuss research related to our work. As our main contribution, we propose in section 4 a new selection scheme that is able to gather a small and robust set of local features. Based on our approach the mentioned shortcomings of local features are alleviated, which we finally prove in section 5 by empirical results on two different datasets.

2 DISCRIMINATIVE LOCAL FEATURES FOR TRAFFIC SIGN RECOGNITION

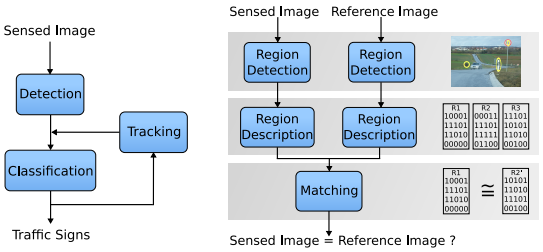
Traditionally, the task of TSR is divided into 3 stages: i) TSD, the detection of traffic sign candidates, ii) TSC, the classification of the candidate to the proper traffic sign class (i.e. the type of the road sign, e.g. a specific speed limit), and iii) the tracking of the candidate within the video sequence (cf. Fig. 1(a)).

Local features can be applied to all three stages of TSR. Since our considerations hold for every stage, we do not focus on a single one in this paper. The same applies to the choice of a specific local feature method. Although, the proposed selection approach is valid for several methods, for the evaluation of our approach we restrict ourselves to SIFT, since we identified SIFT to be most suitable for TSR. We make this decision based on the evaluation of feature descriptors provided by Mikolajczyk and Schmid [MS05]. In their evaluation, SIFT pointed out to be the strongest local feature in terms of recall and precision of all tested approaches. Additionally, we tested some recent local feature approaches for their suitability to TSR, among them SIFT [Low04], GLOH (gradient location and orientation histogram) [MS05] and some variants of SURF [BTG06]. The results are summarized in Table 1 and show the matching performance of these local feature approaches between sensed traffic sign images and traffic sign features stored in a database. This evaluation was done on a challenging testset (degraded traffic signs at low resolution) of 46 images with 99 traffic signs extracted from a 30 minute video sequence, captured while driving on a highway. Note that we do not provide a complete evaluation of local feature methods also considering more discriminative color versions of the mentioned techniques, since this is out of scope for this paper and does not affect our selection approach.

The common fashion, discriminative local features are used, is depicted in the 3 stages of Fig. 1(b). Basically the descriptors of both images, the sensed and the reference image, are calculated and compared to each other. Often the descriptors of the reference image are previously extracted and stored in a database. Note that we derive the terminology of sensed and reference image from the field of image registration. In the context of TSR, *sensed images* are the images of the traffic scene, which are commonly gathered by a front-view camera behind the car's windshield, while the term *reference image* refers to the image of which the descriptors are extracted from that define the traffic sign class.

	Recall	Precision
SIFT	33.68 %	78.05 %
GLOH	22.11 %	52.50 %
SURF	7.37 %	35.00 %
USURF128	25.26 %	72.73 %

Table 1: Performance of recent local feature approaches in the context of TSR.



(a) (b)

Figure 1: (a) Structure of TSR. (b) Stages of local feature matching.

As illustrated in Fig. 1(b), the matching process of two images can be divided into 3 parts. We now have a closer look into these parts and point out the specific issues arising with the application of discriminative local features in the domain of TSR.

First of all, an interest point detector is used to extract interest points that are likely retrievable. In the case of SIFT, this detector searches for extrema within the difference-of-gaussian pyramid, an approximation of the laplacian of the $2+1$D scale-space. In common, these regions are very salient and thus likely to be retrieved. An issue arises with the region's retrievability under different image scales. For example regions around coarse image structures can be retrieved from images with low resolution, while others cannot. This is due to the loss of detail caused by decreasing spatial resolution or increasing distance from the captured object (e.g. the traffic sign). Furthermore, salient regions may appear at certain scales of an image due to sampling artifacts. Feature descriptors calculated from these regions are often ambiguous and weakly discriminative and may cause false positive matches (*FP*) (cf. Fig. 5(e)). Hence, robustness to the image scale is important. Besides this, the regions extracted by an interest point detector have to be located at corners to be positionally stable, especially if the sensed image is captured from another perspective. In the traffic sign domain the positional stability is challenging, since traffic signs often include circular symbols, so that corners are

quite rare. This leads to unstable regions' locations and thus to feature descriptors that do not match with the corresponding features of the reference sign.

The second step in Fig. 1(b) depicts the descriptor calculation, which is performed after region detection. In the case of SIFT, the region patch is expressed by a 128 dimensional feature vector that is calculated from the region's edge orientation histogram. These features are often salient, that is rare in feature space, if they are calculated on a textured region. Unfortunately, in the domain of traffic sign recognition we have to deal with highly structured objects, which means the regions only include a few, very basic geometrical shapes. This leads to features with weak discriminative power. These features are unsuitable for TSR, since the following two problems may occur. If the sensed image does not contain the according object of the reference image, these features are likely to be confused with other features, resulting in false positive matches. The other problem is the inhibition of correct matches depending on the similarity measure used. Also challenging are the feature descriptions of regions that violate the planarity assumption, since they are viewpoint dependent. In the context of TSR, this issue is reduced to regions that overlap the traffic sign border, so that their features partially describe the background. This leads to features varying with the background captured around the traffic sign.

Finally, the feature descriptors of the reference image are compared with those extracted from the sensed image, as illustrated in the last stage of Fig. 1(b). This introduces a similarity measure to distinguish the features. Among the Mahalanobis distance, the Euclidean distance is very popular and was used for object recognition with SIFT by Lowe [Low04]. Lowe proposes a matching scheme with distance ratio $T_{ratio} = 0.8$ to the second nearest neighbor. So, matches have to satisfy

$$\frac{|\mathbf{v} - \mathbf{w_1}|}{|\mathbf{v} - \mathbf{w_2}|} < T_{ratio}$$

where $\mathbf{w_1}$ is the nearest neighbor and $\mathbf{w_2}$ the second nearest neighbor to feature vector $\mathbf{v}$. We compare every feature of the sensed image with those extracted from the reference sign (or stored in a database). Applying this order we receive a higher recall, than by matching vice versa. Additionally, this matching direction can more easily benefit from approaches for approximate nearest neighbor retrieval like *Best Bin First* [BL97].

A large number of FP may occur if this matching direction is used with a small number of features representing the reference image. This originates from the sparse feature space and can be alleviated by the introduction of an additional a distance threshold. We identified a distance threshold of $T_{dist} = 0.425$ to suit best the needs of TSR. This threshold is derived as 30 % of the maximal distance ($\sqrt{2}$) in the normalized feature space

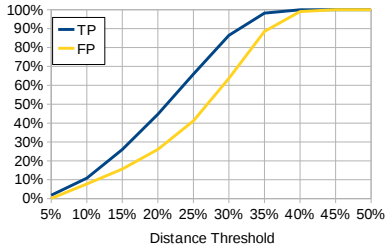


Figure 2: Matching performance in TP and FP (compared to TP and FP using only distance ratio matching) with additional distance threshold.

of $\mathbb{R}^d_+$ ($d > 1$) using the euclidean metric. Fig. 2 illustrates the dependency of true positive matches (*TP*) and FP according to the distance threshold T_{dist} obtained on a subset of the "Affine Covariant Regions Datasets" [MTS+05]. The problems arising in the matching stage are not restricted to the traffic sign domain only, but are rather common for most local feature applications. The main issue is the time spent for feature comparison, especially when the reference feature database is large. This is due to the search of the nearest neighbor for each feature vector of the sensed image. Hence, the number of features stored in the database for each traffic sign class is an important factor for the duration of the feature comparison. This number directly depends on the reference image's size and the complexity of the content as it is depicted in Fig. 3 for three different traffic signs. Thus, the issue is to select the right resolution of the reference image to receive a manageable number of features: not too many, due to speed reasons, but also not too few in order to safely recognize the traffic signs of the sensed image, even if the image is captured under severe conditions or is partially occluded.

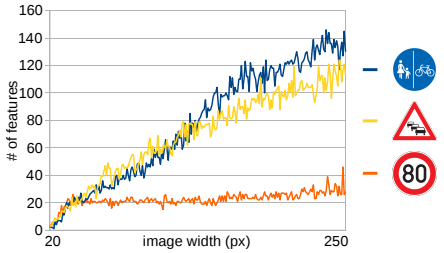


Figure 3: The number of features extracted from a reference image depends on the content as well as on its resolution.

3 RELATED WORK

Now that we have depicted the problems arising with the application of discriminative local features in the traffic sign domain, we give an overview of the research in this field. Recently, local features are utilized in the context of TSR in different ways. They are applied to the detection of traffic signs as well as to their classification.

In the field of TSD, weakly discriminative rectangle features in combination with a boosted classifier cascade are used to detect potential sign candidates [BV04, ERP07, BEV+09]. In the works of Bahlmann *et al.* [BZR+05] and Keller *et al.* [KSB+08] this technique is extended and the Haar-like features are calculated on 7 different chromatic bands to increase the discriminative power of these approaches. In contrast to the mentioned approaches, Höferlin and Zimmermann [HZ09] apply highly discriminative SIFT features to the problem of TSD. In their approach, a subsequent classification stage benefits from the estimation of the traffic sign class of the preceding SIFT detector. Contrary to this, Kus *et al.* [KGEU08] apply SIFT in combination with color features to the detection as well as to the classification of traffic signs.

Discriminative local features are adopted by Farag and Abdel-Hakim [FAH04] to classify traffic signs. They use SIFT features to classify the previously detected candidates. Ruta *et al.* [RLL07] introduce a distance transform based on color and use the resulting image representation for their local feature selection. Their approach for TSC was inspired by the trainable similarity measure used by Paclík *et al.* [PND06].

Local features are also applied to object tracking in video sequences. This topic is covered by the survey of Trucco and Plakas [TP06], to which we refer.

To reduce the amount of time spent in feature matching two methods are possible. First: the retrieval of the nearest neighbor can be accelerated. And second: the number of features can be reduced. For the first option a large number of methods already exist in literature. The problem is known as *Nearest Neighbor Search* and algorithms based on data structures like the kd-tree exist, which retrieves the nearest neighbor averagely in logarithmic time complexity for low dimensional spaces. But it is shown [IM98], that with the number of dimensions the nearest neighbor search approaches the expense of an exhaustive search. This issue is called the *Curse of Dimensionality* and can be alleviated by easing the restrictions and by avoiding the assurance of retrieving the exact nearest neighbor. The altered problem is called the *Approximate Nearest Neighbor Search*. There are also well-known methods to solve this problem like *Best Bin First* [BL97], which in most cases finds the nearest neighbor and otherwise retrieves another close candidate.

The second option is covered by Joly and Buisson, they extract Harris interest points from video data and consider only those features for matching, that are salient among all features in their database [JB05]. This way they speed-up the comparison process and obtain features of high quality. Their work inspired our method for the selection of discriminative features. They define the saliency of a feature according to Walker *et al.* [WCT98] as the likelihood of being misclassified with another feature. This definition leads towards density estimation in feature space to receive the probability density function of feature misclassification. In contrast to Joly and Buisson we also consider the robustness and the retrievability of features, that increases the matching performance, too.

Viewpoint invariance of local features is only little covered by research so far, e.g. [VS06].

4 SELECTION OF FEATURES

Based on the application of local features in TSR in section 2, we develop in this section a method for the selection of a set of robust and discriminative local features. We choose these features optimally with respect to the mentioned issues.

4.1 Retrievability and Localization Stability under Scale Changes

The retrievability as well as the positional stability of scale-adapted interest regions vary with the image scale and depend on the image structure that is covered by the region. Interest regions that can be retrieved under multiple scales are mostly located at coarse image structures (i.e. areas of low detail or low image frequency). The idea is to select those features which show the highest robustness and retrievability under scaling. We measure the retrievability of a feature by counting its occurrences at several scales. For every resolution we extract the features and match them to the features of the previous scales. Matching is done by searching for the nearest neighbor in a combined distance space of the feature vector and the keypoint's position, scale, and orientation. The distances are euclidean and both are separately thresholded to avoid false matches. If coinciding features are found, the arithmetical mean is calculated for their descriptors and positions, weighted by the number a feature was extracted in the previous scales (Fig. 4). The arithmetical mean centers the feature in image space and in feature space to reduce the influence of the interest point's localization instabilities and the variance of the region's descriptor.

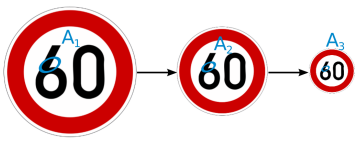


Figure 4: Selection of a region's descriptor that is robust to the change of image resolution.

Further, we measure the localization stability of a interest point under image scale changes. Brown *et al.* show, that common interest point detectors like SIFT lack of accuracy in repeatability of the interest point's

position, scale and orientation [BSW05]. They illustrate that such detectors extract the majority of their interest points with a positional variation between 0 and 3 pixels. In our tests, we experienced in average a jitter of about 0.4 % of the interest point's position, according to the image resolution. This corresponds to an expectation value of about 1 pixel based on the considered resolution range of 20 to 500 pixels in width. There is also a strong relation between the tapering of corners in the image and the accuracy of interest point localization. Hence, the average of the positional accuracy for traffic signs is about three times worse than for the natural images that we tested, although the maximal standard deviation of the interest point's position is quite similar among all tested examples (about 3 %). A similar effect is observed for the accuracy of scale and orientation. The regions' scales calculated by SIFT are disturbed in average by 0.2 % and we observed a maximal standard deviation of 0.5 % (0.05 %, 2.44 % for orientation).

Since the positional accuracy of interest points depends on the underlying image structure and varies between interest points, we also rate their quality in order to select a robust feature set. We derive the rating score from the standard deviation of the interest point's position, scale, and orientation, which is calculated on scaled instances of the reference image. Fig. 5(a) and (b) show the 5 best and the 5 poorest features according to their retrievability and positional stability under image's scale changes. As expected, features that are well retrievable among different image scales do more likely cover regions of coarse image structures. In contrast regions with poor retrievability are located at structures that show higher frequencies. Regions with high stability of position, scale, and orientation are small in size and generally extracted at tapering corners, while the regions of poor stability are located at less distinct corners (see Fig. 5(b)).

4.2 Viewpoint Variant Features

Recent discriminative local feature approaches like SIFT are invariant to rotation, scale and to some degree to affine deformations. In contrast to this, the most of these approaches are sensitive to the change of viewpoint under which the image is captured. This is especially a problem if the image patch of the detected interest point strongly violates the local planarity assumption, which is the case at ridges, corners, and occluding boundaries. Vedaldi and Soatto [VS06] try to cope with such regions by considering the scene geometry. But since the scene geometry is not always available or could be made available, we head for another method.

In our approach, we avoid additional knowledge of the scene geometry, even if it is simple like in the case of traffic signs. Therefore, we utilize a small set of training images that are captured from different viewpoints in order to measure the variation of the feature's descriptor. This method yields the additional advantage that the variation intensity is considered, too. Thus, violations to the local planarity assumption may be neglected, if they marginally affect the feature vector.

We determine the robustness to viewpoint changes of a feature by considering the spatial distribution of the distances between instances of features within the training set. For example, the subdivision of the feature patch into 16 subregions represents the spatial information of the SIFT descriptor. Thus, we measure the subregions' distance variation σ_{dist} for every instance of a particular descriptor within the training set:

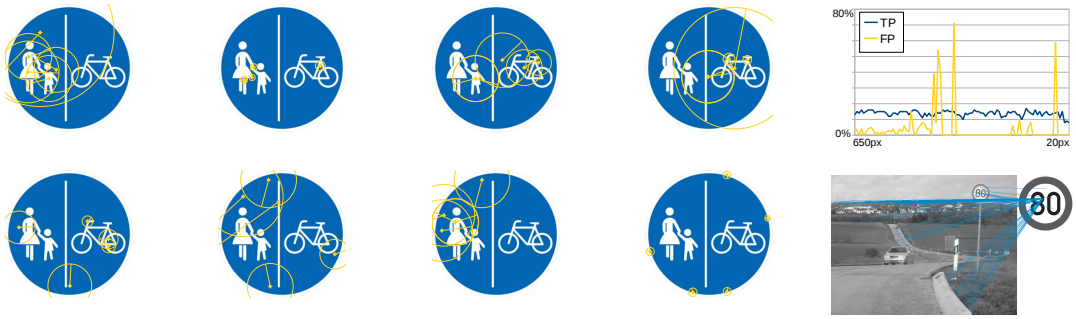
$$\sigma_{dist} = \sqrt{\sum_{k=1}^{r} (D_{k,k} - \hat{E})^2}$$

For the 128-dimensional SIFT descriptor, the distance of regions between two feature instances i and j is defined by $D_{1..r,1..r} = \sqrt[*]{(M_i - M_j)^T (M_i - M_j)}$, where M_i and M_j are their reshaped 8×16 feature matrices. The component-wise square root function is $\sqrt[*]{\cdot}$ and $\hat{E} = tr(D)/r$ represents the mean distance of the $r = 16$ subregions. The idea behind this concept is that patches with small distance variations are in common just jittered or misaligned, whereas large distance variations often indicate the inclusion of ridges, corners, and object boundaries. We use the maximum standard deviation $max(\sigma_{dist})$ of all instances of a single feature within the training set as measure for the region's stability to viewpoint changes. In the case of traffic signs, only occluding boundaries affect the region's stability with the change of viewpoint or background. Hence, stable region patches may only inclose the traffic sign area, while patches with poor stability to viewpoint changes overlap the sign boundaries (c.f. Fig. 5(c)).

4.3 Analyzing the Discriminative Power

Generally, it is desirable to select a set of highly discriminative features. This raises the probability for correct matches and reduces the likelihood of feature vectors to constrain each other's matching performance. During the evaluation of interest point detectors, we noticed the occurrence of features with marginal discriminative power at certain image sizes, originating from sampling artifacts. In Fig. 5(e) the matching performance of SIFT features at different scales is visualized. It points out that certain scales show sudden peaks of FP. These peaks originate from highly indiscriminative features as depicted in the bottom row of Fig. 5(e).

The discriminative power of a feature vector can be expressed by its saliency, which is equivalent to its rarity in feature space. Therefore, the probability density function of retrieving the wrong nearest neighbor is directly correlated with the density of the feature's neighborhood. This introduces kernel density estimation as



(a) (b) (c) (d) (e)

Figure 5: The 5 best rated features (circles, top row), the 5 poorest regions (circles, bottom row). Both depicted on an exemplary reference sign. (a) Retrievability under scale changes, (b) positional stability, (c) descriptor stability on viewpoint changes and (d) saliency of reference features among the features extracted from MIT-Highway set. Further explanation in the text. (e) Features with low discriminative power. Top: occurrence of these features at some scales. Bottom: interest point relation of these features.

appropriate measure for the discriminative power of a feature. We estimate the density function $\hat{f}$ at point v using a Gaussian kernel by

$$\hat{f}(v) = \frac{1}{|M|} \frac{1}{\sigma(2\pi)^{\frac{d}{2}}} \sum_{x \in \xi} e^{\frac{\|v-x\|^2}{2\sigma^2}} \qquad (1)$$

The set of features is represented by ξ and d refers to the dimensionality of the feature space which is 128 in the case of SIFT descriptors. We choose bandwith $\sigma = 0.3\sqrt{2}$ to be equal to the distance threshold T_{dist}. That corresponds to a window radius of 30% of the maximum distance in normalized feature space. We populate the feature space ξ with features from typical traffic scene images, harvested from the MIT-Highway set. The bottom row of Fig. 5(d) shows an example of features with poor saliency. These are located along the traffic sign border and do not contain any discriminative image structures, while the top row regions include salient structures.

Besides feature's domain specific saliency, there are two other groups of features that are important in order to obtain a high matching performance. First, ambiguous features within the reference image should be avoided, since similar descriptors tend to inhibit each other, when nearest neighbor matching with distance ratio is used. Same holds for the second case: the saliency of features across the traffic sign classes. Obviously, a higher recall performance and also a better distinction between the sign classes can be achieved if features are selected that are discriminative across all classes. For both cases, we also apply the Parzen window approach according to 1, but either ξ is populated with the features of one traffic sign class or with the descriptors of all other traffic sign classes.

4.4 Choosing the Optimal Feature Set

Now that we have introduced a variety of criteria for robust and discriminative features, we select an optimal set of n features to represent the traffic sign class. We define the overall quality q of feature f by a weighted sum of its scores achieved for each criterion c_i:

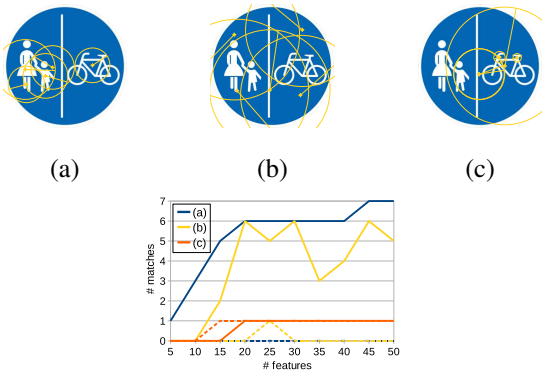
$$q(f,S) = \sum_i w_i c_i(f,S)$$

The feature set can be adapted to a particular problem by adjusting the weights w_i of each criterion. Since the quality q of a feature depends on the other features selected for the representation set S, we have to solve a non-linear optimization problem. There are various approaches to approximate the solution of a non-linear optimization problem, e.g. evolutionary algorithms. We use a greedy algorithm to choose in every iteration step the feature with the highest fitness quality q, with respect to the features S selected in previous steps. For initialization, we start with an empty set S. This simple approach is very fast and yields good results, even if the detection of the global optimum is not ensured and thus the set may only be locally optimal.

5 EVALUATION

For the evaluation of the proposed method we first compare the feature sets obtained by several feature selection approaches for an exemplary traffic sign. Then, we compare the performance of our approach with that of the conventional representation of the traffic sign class as reference image of a certain size.

Fig. 6 shows the sets of 5 features selected by three different approaches: (a) the proposed approach, (b) selection by choosing a proper image resolution, and (c) selection of the most discriminative features, similar to Joly and Buisson [JB05]. Fig. 6(a) points out that the features retrieved by the proposed method are located on discriminative and coarse image structures, while the other both approaches either lack of the one or the other property. In Fig. 6(d) the matching performance of the 3 approaches is compared for the exemplary traffic sign class. It points out that especially



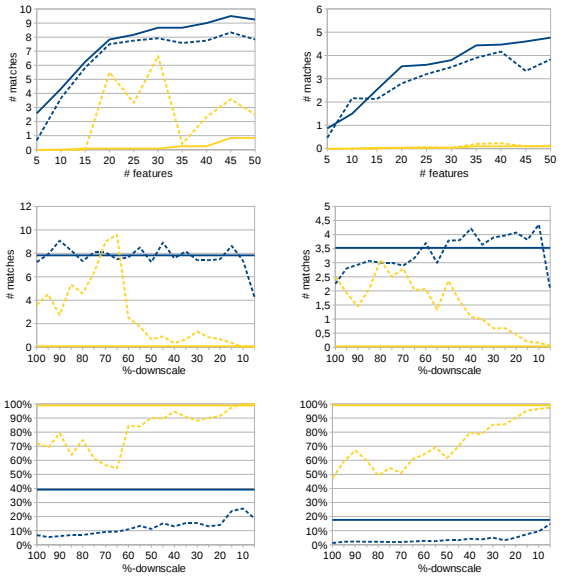
(d)

Figure 6: Exemplary feature set on a reference sign. Only 5 features shown for clarity. (a) Proposed method, (b) extracted for a certain reference image resolution, (c) selected at low density in feature space and (d) performance of these methods for increasing number of features. TP (solid line), FP (dashed line).

the features retrieved by method (c) have poor matching performance, since the most discriminative features of the reference sign often show low retrievability under scale changes.

We use two different testsets for further evaluation. The first contains 12 images (640x480 px) with 9 traffic sign classes, among them images captured under adverse conditions: blur, noise, perspective deformation and chromatic lighting. We call it the "German" testset, since it covers German traffic signs. The second set includes 30 images (360x270 px) of 3 classes. It is introduced in [GP03]. We refer to it as the "Dutch" set for the same reason. We compare the matching performance of the features selected with the proposed method (a set of training images per sign class) to those extracted from reference images of a certain size, as it is the conventional method to limit the number of features. That means, the selection parameter is the image resolution. For evaluation we restrict the number of selected features to 20. The results of 3 experiments for both testsets are presented in Fig. 7. The performance measure is provided by means of TP and FP.

The first experiment (Fig. 7(top row)) shows the matching performance of both methods with increasing feature count. The reference image's size was adapted in case of the conventional method in order to provide the proper number of features. Obviously, the matching performance benefits from the features selected by the proposed method. Especially, the low number of false positives for the "German" testset is remarkable, while the conventional method suffers from indiscriminative features at some resolutions. For the chart presented in the middle row of Fig. 7 we do not longer fix the number of features selected by the conventional method (for the proposed method we use a set of 20 features) but show the performance of the features



(a) (b)

Figure 7: Average matching performance. Top and middle row: TP in blue/dark gray, FP in yellow/light gray. Bottom row: Recall in blue/dark gray, precision in yellow/light gray. Proposed method (solid line), conventional method (dashed line). (a) "German" testset, (b) "Dutch" testset.

extracted at different scales. We see that it is difficult to choose a certain reference image size with high TP and low FP. This becomes even more obvious if we involve the number of features and thus express the performance by means of recall and precision (c.f. Fig. 7(bottom row)). Our method also reduces the time spent for matching, since the desired number of features can be defined by the user (cp. Fig. 3). Hence, feature dependencies to the reference image's resolution are eliminated and the matching duration becomes appreciable.

6 CONCLUSION

We presented a method to select a small set of discriminative local features (e.g. SIFT features) with excellent matching properties. Applying our approach we were able to increase the matching performance by speed, recall and precision. Hence, issues inherent to local feature matching were alleviated. Further work has to consider other selection approaches to retrieve the globally optimal set of features. In addition, new criteria like the spatial distance of the interest points or the coverage of the reference object by the set of features can be introduced, to increase robustness against occlusion, too.

REFERENCES

[BEV+09] X. Baró, S. Escalera, J. Vitrià, O. Pujol, and P. Radeva. Traffic Sign Recognition Using Evolutionary Adaboost Detection

and Forest-ECOC Classification. *IEEE Transactions on Intelligent Transportation Systems*, 10(1):113–126, 2009.

[BL97] JS Beis and DG Lowe. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *Proceedings to IEEE CVPR'97.*, pages 1000–1006, 1997.

[BL07] M. Brown and D.G. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73, 2007.

[BSW05] M. Brown, R. Szeliski, and S. Winder. Multi-image matching using multi-scale oriented patches. *IEEE CVPR 2005.*, 1:510–517 vol. 1, 2005.

[BT05] VTT Building and Transport. Analysis of context of use and definition of critical scenarios, 2005. EU project HUMANIST. Reference: AVTT-030305-T1-DA(1).

[BTG06] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. *SURF: Speeded Up Robust Features*, volume 3951 of *Lecture Notes in Computer Science*, chapter Computer Vision - ECCV 2006, pages 404–417. Springer Berlin / Heidelberg, 2006.

[BV04] X. Baró and J. Vitrià. Fast traffic sign detection on greyscale images. *Recent Advances in Artificial Intelligence Research and Development*, pages 69–76, 2004.

[BZR+05] C. Bahlmann, Y. Zhu, Visvanathan Ramesh, M. Pellkofer, and T. Koehler. A system for traffic sign detection, tracking, and recognition using color, shape, and motion information. *IEEE Intelligent Vehicles Symposium.*, 2005.

[ERP07] Sergio Escalera, Petia Radeva, and Oriol Pujol. Traffic sign classification using error correcting techniques. *VISAPP 2007*, pages 281–289, 2007.

[FAH04] Aly A. Farag and Alaa E. Abdel-Hakim. Detection, categorization and recognition of road signs for autonomous navigation. *Proceedings of Acivs 2004*, pages 125–130, 2004.

[GP03] C. Grigorescu and N. Petkov. Distance sets for shape filters and shape recognition. *IEEE Transactions on Image Processing*, 12(10):1274–1286, 2003.

[HZ09] B. Höferlin and K. Zimmermann. Towards reliable traffic sign recognition. In *Intelligent Vehicles Symposium, 2009 IEEE*, pages 324–329, June 2009.

[IM98] P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *18th ACM Symposium on Theory of Computing*, pages 604–613. ACM New York, NY, USA, 1998.

[JB05] Alexis Joly and Oliver Buisson. Discriminant local features selection using efficient density estimation in a large database. In *Proceedings to 7th ACM SIGMM MIR*, pages 201–208, 2005.

[KGEU08] M.C. Kus, M. Gokmen, and S. Etaner-Uyar. Traffic sign recognition using Scale Invariant Feature Transform and color classification. In *Computer and Information Sciences, 2008. ISCIS'08. 23rd International Symposium on*, pages 1–6, 2008.

[KSB+08] Christoph Gustav Keller, Christoph Sprunk, Claus Bahlmann, Jan Giebel, and Gregory Baratoff. Real-time recognition of u.s. speed signs. In *IEEE IV'08*, 2008.

[Low04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[MS05] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(10), 2005.

[MTS+05] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *Int. J. Comput. Vision*, 65(1-2):43–72, 2005.

[PND06] P. Paclík, J. Novovicova, and RPW Duin. Building road-sign classifiers using a trainable similarity measure. *IEEE Transactions on Intelligent Transportation Systems*, 7(3):309–321, 2006.

[RLL07] A. Ruta, Y. Li, and X. Liu. Traffic sign recognition using discriminative local features. *Lecture Notes in Computer Science*, 4723:355, 2007.

[TP06] E. Trucco and K. Plakas. Video tracking: a concise survey. *IEEE Journal of Oceanic Engineering*, 31(2):520–529, 2006.

[VS06] A. Vedaldi and S. Soatto. Viewpoint induced deformation statistics and the design of viewpoint invariant features: Singularities and occlusions. *Lecture Notes In Computer Science*, 3952:360, 2006.

[WCT98] KN Walker, TF Cootes, and CJ Taylor. Locating salient object features. In *BMVC'98*, pages 557–566, 1998.