

Building Road-Sign Classifiers Using a Trainable Similarity Measure

Pavel Paclík, Jana Novovičová, and Robert P. W. Duin, *Member, IEEE*

Abstract—Deriving an informative data representation is an important prerequisite when designing road-sign classifiers. A frequently used strategy for road-sign classification is based on the normalized cross correlation similarity to class prototypes followed by the nearest neighbor classifier. Because of the global nature of the cross correlation similarity, this method suffers from presence of uninformative pixels (caused, e.g., by occlusions) and is computationally demanding. In this paper, a novel concept of a trainable similarity measure is introduced, which alleviates these shortcomings. The similarity is based on individual matches in a set of local image regions. The set of regions that are relevant for a particular similarity assessment is refined by the training process. It is illustrated on a set of experiments with road-sign-classification problems that the trainable similarity yields high-performance data representations and classifiers. Apart from a multiclass classification accuracy, nonsign rejection capability and computational demands in execution are also discussed. It appears that the trainable similarity representation alleviates some difficulties of other algorithms that are currently used in road-sign classification.

Index Terms—Classifier system design, road-sign classification, similarity data representation.

I. INTRODUCTION

THE road-sign-recognition system is a module of a driver support system of an intelligent vehicle. It should contribute to updating a world model with information on traffic signs present in the current traffic situation. Although road signs are man-made objects defined by international standards, a number of issues turn the automatic road-sign recognition into a challenging problem. Let us name at least the large number of signs types (classes) to be distinguished, numerous country- or vendor-specific sign variants [1], general illumination conditions, or sensor vibrations. In order to design a system (resilient to noisy conditions), a statistical pattern recognition approach is usually adopted [2]–[4]. Based on a database of labeled examples, a road-sign-recognition system is trained, minimizing the error expected on examples unseen in training. Apart from high

accuracy in classification of different sign types, a recognition system should also avoid erroneous identification of nonsigns, i.e., limit the number of false alarms. Furthermore, it should be suited for real-time deployment.

When the road-sign-recognition problem is considered in full generality, the sign detection and classification stages are usually distinguished [4]–[7]. In a traffic scene image, a road-sign detector identifies a set of candidate regions using, e.g., edge-based template matching [8] or color segmentation algorithms [9]. Each candidate region is then passed on to a classification module and either assigned to one of the known road-sign classes or rejected as a nonsign. In this paper, we adopt the separation between detection and classification and focus on the design of the classification module.

In order to design a road-sign classifier, the input, variable sized, candidate regions must be appropriately represented for the given classification technique. After a basic preprocessing such as scaling of the regions to equal size or masking out the general sign background (BG), a more specific data representation has to be constructed. So far, two conceptually different data representations have been used for road-sign classification, namely the feature-based and similarity-based approach.

In the first case, each candidate region is represented by a vector of numerical characteristics (features). Examples of features used in road-sign classification are color histograms [9], moment invariants, and *ad hoc* image characteristics [1], [10], wavelets [6], [11], appearance-based features [12], [13], or, directly, the subsampled pixel intensities [14]–[17].

In a similarity-based approach, each candidate region is represented by a set of similarities to stored prototype examples. The representation is, therefore, relative, contrary to the absolute description of a patch by a feature vector. The advantage of a relative representation is that the application-specific evidence is supplied directly by prototype objects and not refined from a set of general-purpose features. The similarity-based representation may also account for different modes in the data by utilizing corresponding prototype objects. This may be especially advantageous for road-sign classes composed of subgroups (e.g., speed limits). In this paper, we focus on the construction of the similarity-based data representation.

The most frequently used similarity measure in road-sign classification is a normalized cross correlation [5], [18], [19]. It is advantageous due to its simplicity, robustness to varying illumination conditions, and statistical interpretation [20], [21]. It is also appealing from the implementation viewpoint, as the optical correlator for road-sign recognition has been successfully demonstrated by Guibert *et al.* [22]. However, as a global measure, the normalized cross correlation is sensitive

Manuscript received October 27, 2004; revised July 5, 2005, December 19, 2005, February 16, 2006, and March 18, 2006. This work was supported by the Grant Agency of the CAS under Grant A2075302, the Complex Research Project of the CAS under Grant K1019101, and the Grant Agency of the Czech Republic under Grant 102/03/0049. The Associate Editor for this paper was N. Zheng.

P. Paclík and R. P. W. Duin are with the Information and Communication Theory Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: P.Paclik@ewi.tudelft.nl; R.Duin@ieee.org).

J. Novovičová is with the Department of Pattern Recognition, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, 182 08 Prague 8, Czech Republic, and also with the Faculty of Transportation Sciences, Czech Technical University, Prague, Czech Republic (e-mail: novovic@utia.cas.cz).

Digital Object Identifier 10.1109/TITS.2006.880627

to the presence of uninformative pixels and therefore suffers from template misalignments. Processing the full size candidate regions in each comparison is also computationally expensive.

The goal of this research is to develop a similarity measure which could be specifically tuned to a particular road-sign-classification problem, be more resilient to uninformative pixels, and be computationally less expensive than the normalized cross correlation. Our proposal is to base a similarity assessment only on local clues relevant in a given comparison. Those, naturally, depend on the used prototype object. The desirable local measure should be, therefore, asymmetric and prototype-specific, unlike the currently used symmetric similarities.

The original contribution of this paper is the proposal to train such a local similarity measure from examples. The directional similarity of an arbitrary image to a particular prototype is computed from a set of matches in local image regions. The set of regions, relevant when measuring a similarity to a given prototype, is refined by the training process (thereby, we refer to the measure as the trainable similarity). The training maximizes separability of the class of the prototype object from the remaining classes. The class membership of prototype objects is readily available because prototype objects are usually selected from the labeled training set.

Our approach is close to the work of Hsu and Huang [6] who used the matching-pursuit approach for road-sign classification. Matching pursuit is a greedy algorithm for the determination of an overcomplete set of bases using localized wavelet filters. The authors propose a two-step process. First, assuming that classes are compact, the set of bases is identified. In the second stage, each class is described by a subset of bases providing the highest discrimination with respect to the remaining classes. For a new image to be classified, a coefficient (feature) vector is computed for each class using a class-specific set of bases. The coefficient vector is then compared to the stored class centroid using the cosine distance and assigned to the closest class in a nearest neighbor fashion.

The proposed trainable similarity measure differs from the matching-pursuit approach in several important aspects. While the method of Hsu and Huang constructs an intermediate feature representation and computes the dissimilarities between feature vectors, our trainable similarity compares an image to the prototype by directly matching the pixel intensities. The matching-pursuit approach uses an assumption of compact classes both when deriving a suitable set of bases and in classification (each class is represented by its centroid). In this paper, we illustrate that the concept of a trainable similarity is applicable to multimodal situations using diverse prototype selection strategies. Finally, the number of extracted bases in the approach of Hsu and Huang needs to be specified by the user. On the contrary, the number of local regions used in assessing a similarity to a particular prototype is identified automatically by the training procedure.

Firstly, the outline of the road-sign-classification problem is given and the correlation similarity measures are introduced. Then, the new trainable similarity measure is discussed together with the region selection strategy and the corresponding criterion utilized in training. Possible approaches for building a road-sign classifier based on the trainable similarity are

outlined in Section IV. In Section V, the construction and behavior of the trainable similarity are illustrated on a set of experiments. The performance comparison of classifiers based on the trainable similarity representation with several other road-sign-classification strategies is given in Section VI. Apart from multiclass classification accuracy, rejection capability and classifier speed are also evaluated. Finally, Section VII presents the conclusions.

II. SIMILARITY BASED REPRESENTATION OF IMAGES

In this paper, we assume that the preprocessing converts variable sized input regions \mathcal{I}_j into images I_j of equal size of $n \times n$ pixels, $j = 1, 2, \dots, N$. Let $S(I, J)$ be a measure of similarity between two images I and J . Let $\{Pr_1, \dots, Pr_{N_p}\}$ be a given set of N_p labeled prototype images. We introduce a representation of an image I by using a set of similarities $\{S(I, Pr_1), S(I, Pr_2), \dots, S(I, Pr_{N_p})\}$. Thus, an image I can be represented by N_p -dimensional vector $(S(I, Pr_1), S(I, Pr_2), \dots, S(I, Pr_{N_p}))$.

A. Correlation-Based Similarity Measures

Let I^i and J^i represent intensities of i th pixel of corresponding images I and J , $i = 1, 2, \dots, n$. The correlation-based measures of similarity between two images I and J of equal size are presented, e.g., in [20] and [21]. The cross correlation between two images I and J represented by the intensities is their inner product $S_R(I, J) = \sum_i I^i J^i$. The normalized cross correlation that is invariant to scaling of pixel intensities is a more convenient measure for image matching

$$S_{Rn}(I, J) = \frac{\sum_i I^i J^i}{\sqrt{\sum_i (I^i)^2 \sum_i (J^i)^2}}. \quad (1)$$

The cross correlation coefficient $S_r(I, J)$ between two images I and J is defined as

$$S_r(I, J) = \frac{\sum_i (I^i - \bar{I})(J^i - \bar{J})}{\sqrt{\sum_i (I^i - \bar{I})^2 \sum_i (J^i - \bar{J})^2}} \quad (2)$$

where \bar{I} and \bar{J} denote the corresponding means of image intensities. The measure $S_r(I, J)$ appraises the degree of linear dependence between two images being compared and has the following properties.

- 1) $S_r(I, J) \in \langle -1; 1 \rangle$. A value of one indicates perfect matching (identical images, except for scale and offset), and zero corresponds to complete mismatch and a value of minus one to the perfect negative match (e.g., one image is a negative of the other).
- 2) $S_r(I, J)$ is invariant to linear scaling and shift of intensity values between I and J .
- 3) $S_r(I, J)$ is symmetrical: $S_r(I, J) = S_r(J, I)$.
- 4) $S_r(I, J)$ is nonrobust in that a single outlying pixel can distort them arbitrarily.
- 5) $S_r(I, J)$ is not suitable in the presence of nonlinear intensity variation at corresponding pixels.

III. TRAINABLE SIMILARITY MEASURE FOR IMAGES

In order to define a trainable similarity measure, let $\Omega = \{\omega_1, \dots, \omega_C\}$ be a set of predefined classes and $Tr = \{(I_1, \omega(I_1)), \dots, (I_N, \omega(I_N))\}$ a set of N labeled training variable size input regions.

The similarity $S(I, J, R)$ between equal-sized rectangular images I and J is based on local image matches in a set of local image regions R . Each local region $r \in R$ is defined by its corner coordinates: $r = \{x_1, y_1, x_2, y_2\}$. The local match $s(I, J, r)$ between the corresponding pixel values $\mathbf{r}(I)$ and $\mathbf{r}(J)$ may be measured, for example, by a normalized cross correlation

$$s(I, J, r) = \frac{\langle \mathbf{r}(I), \mathbf{r}(J) \rangle}{\|\mathbf{r}(I)\| \cdot \|\mathbf{r}(J)\|}. \quad (3)$$

The overall similarity between images I and J is a function of a set of local matches $R = \{r_1, \dots, r_{N_r}\}$

$$S(I, J, R) = f(s(I, J, r_1), \dots, s(I, J, r_{N_r})). \quad (4)$$

Using the mean of local similarities, we obtain the S_{mean} measure

$$S_{\text{mean}}(I, J, R) = \frac{1}{N_r} \sum_{j=1}^{N_r} s(I, J, r_j). \quad (5)$$

Alternative definitions such as S_{min} utilizing minimum of local matches or its robust variant based on lower quartile are discussed in [23].

We assume that deriving the similarity from a set of local matches makes the measure more robust to local image differences caused, for example, by dirt or partial sign occlusions.

A. Training Procedure

The measure presented in this section, calculates the similarity between two images using cross correlation computed in a set of local regions R . The main contribution of this paper lays in the realization that the set of regions R may be tuned to a particular prototype taking into account its class membership. The similarity measure emphasizes only the local clues relevant in the comparison. In order to find the relevant local regions, a training procedure is employed. Based on information derived from a set of training examples, the similarity measure to a particular prototype is refined.

The training procedure requires a labeled prototype image Pr and a labeled set of training images $Tr = \{(I_1, \omega(I_1)), \dots, (I_N, \omega(I_N))\}$. The objective is to find a set R for which the similarities of training examples to the prototype yield the best class separability. We consider here a criterion which emphasizes the differences between the class of a prototype image $\omega(Pr)$ and all other classes in the training set.

Let us first define the set of target and nontarget training images denoting them T and NT , respectively. The target set of the training images includes all images belonging to the same

class as the prototype Pr . The nontarget set includes all other training images

$$\begin{aligned} T &= \{I_i \in Tr : \omega(I_i) = \omega(Pr)\} \\ NT &= \{I_j \in Tr : \omega(I_j) \neq \omega(Pr)\}. \end{aligned} \quad (6)$$

The class of the prototype Pr is denoted in this paper the target class and the collection of other classes from the set Ω the nontarget class.

We search for such a set of regions R for which the target class is the most separable from the non target class. The Fisher's discriminant ratio for two-class problem can be used to quantify the separability capabilities of nontarget images to a prototype

$$\mathcal{C}(Tr, Pr, R) = \frac{(\hat{\mu}_T - \hat{\mu}_{NT})^2}{\hat{\sigma}_T^2 + \hat{\sigma}_{NT}^2} \quad (7)$$

where $\hat{\mu}_T$ and $\hat{\sigma}_T^2$ denote mean and variance of the similarity values $S(I_i, Pr, R)$, $I_i \in T$, respectively. Similarly for $\hat{\mu}_{NT}$ and $\hat{\sigma}_{NT}^2$, respectively. Note that (7) represents a multivariate criterion because a set of local regions is evaluated simultaneously.

In order to find a set R maximizing the criterion $\mathcal{C}(Tr, Pr, R)$, a search strategy is needed. In this paper, we use the sequential forward search strategy (see Algorithm 1) based on the similar idea as the sequential forward feature selection algorithm [24]. Computing the criterion for all singleton regions, the single best region is found and fixed. A new search is launched for an additional region providing the highest criterion value together with the already fixed one. The search continues until all the available regions are used for the construction of a similarity $S(I, J, R)$. Comparison to random search and individual ranking algorithms is discussed in [23].

Note that the eventual number of regions for a particular similarity measure is derived automatically from the training data for each prototype.¹

Algorithm 1 Sequential Forward Selection of Local Regions

- 1: **input:** training set Tr , prototype Pr , region count N_r^{init}
- 2: generate a set of N_r^{init} randomly positioned regions R^{init}
- 3: initialize the pool of unused regions $R^{\text{pool}} = R^{\text{init}}$
- 4: initialize the selected subset $R^{\text{subset}} = \emptyset$
- 5: set the step counter $i = 1$
- 6: while $R^{\text{pool}} \neq \emptyset$ do
- 7: · in a loop over regions $r_k \in R^{\text{pool}}$, $k = 1, \dots, |R^{\text{pool}}|$
- 8: · construct a candidate set $R_k = R^{\text{subset}} \cup r_k$
- 9: · compute the similarity $S(T, Pr, R_k)$ from target examples to the prototype Pr
- 10: · compute the similarity $S(NT, Pr, R_k)$ from nontarget examples to the prototype Pr
- 11: · compute the criterion $\mathcal{C}(Tr, Pr, R_k)$
- 12: · update subset $R^{\text{subset}} = \arg \max_{R_k} \mathcal{C}(Tr, Pr, R_k)$
- 13: · update the pool $R^{\text{pool}} = R^{\text{init}} \setminus R^{\text{subset}}$

¹This is a major difference to an alternative region-selection strategy based on the AdaBoost algorithm, where the number of regions is a user-specified parameter [25], [26].

- 16: · store the best subset $R_i^{\text{subset}} = R^{\text{subset}}$
 17: · store the maximum criterion $\mathcal{C}^{\text{max}}(Tr, Pr, R_i^{\text{subset}})$
 18: · set $i = i + 1$
 19: end
 20: **output:** $R_{\text{best}} = \arg \max_{R_i^{\text{subset}}} \mathcal{C}^{\text{max}}(Tr, Pr, R_i^{\text{subset}})$

IV. CLASSIFIER DESIGN USING SIMILARITY REPRESENTATION

The most commonly adopted strategy for similarity-based road-sign classification is to represent each road-sign class by a single prototype object and then apply the nearest neighbor rule [5], [6], [27]. Formally, the image I , represented by a set of similarities $\{S(I, Pr_j), j = 1, \dots, N_p\}$ is assigned to the class $\omega(Pr_i)$ of the most similar prototype Pr_i

$$Pr_i = \arg \max_{j=1, \dots, N_p} S(I, Pr_j). \quad (8)$$

Several authors noted potential difficulties of this strategy due to its sensitivity to noise. Miura *et al.* [19] proposed to use two thresholds in the nearest neighbor rule. The new-coming observation is assigned to a particular class if the similarity to the corresponding prototype is above the first threshold and if the ratio of similarities to the two closest prototypes surpasses the second threshold. Piccioli *et al.* [5] improve the classification robustness by tracking the candidate regions in a sequence of images and integrating the similarity values.

The performance of the nearest neighbor approach with a single-class prototype also rapidly deteriorates when confronted with a multimodal data distribution or overlapping classes. Although a larger number of prototypes improve the classification accuracy, it also linearly increases the classifier execution time.

The nearest neighbor rule does not utilize the correlations between the computed similarity values. Duin *et al.* recently proposed to leverage the correlations between dissimilarities to prototype objects introducing the concept of a dissimilarity space [28]. Each dimension of such a space measures a dissimilarity to a particular prototype. By projecting the available training observations into the dissimilarity space, one obtains a new data representation where a general-purpose classifier may be built. The same approach was taken independently by Tzomakas for automatic classification of vehicles using ranked-distance correlations [29]. The problem, which is unsolvable by direct ranking, became linearly separable in a space spanned by similarities to prototypes.

Subsequent studies have shown that conventional classifiers built in a dissimilarity space often outperform the full nearest neighbor rule and, in addition, use significantly fewer prototypes [30], [31].

In our recent work, we have illustrated that these classifiers also yield significantly faster execution than the nearest neighbor rule [32]. For a fixed set of prototypes, the parametric classifier built in a dissimilarity space keeps the computational complexity in execution constant. However, its accuracy may still be improved by providing more training examples.

The construction of a similarity space is illustrated by an example in Section V-A.

In order to build a classifier using a similarity data representation, a set of prototype objects must be constructed. Often, ideal noise-free sign images are used as prototypes [5], [27]. In this paper, we advocate the extraction of prototype objects from a training set comprised of real world road-sign examples. The advantage of this approach is that real world prototypes may account for intraclass variations of road signs [1].

Construction of a prototype set may be driven by assumptions on the compactness of a class. A compact class may be sufficiently represented by a single-class prototype. Multimodal classes require, on the other hand, a larger set of prototypes, describing individual class modes. The multimodality is usually caused by the high intraclass variability or broad-class definition (e.g., speed-limit class).

If no assumption on the class structure is made, the selected prototypes should sufficiently cover the problem domain. The best possible coverage is reached if all the available training examples are included in the prototype set, i.e., full set of prototypes. The reduction of the full prototype set is desirable in order to limit the computational complexity. This may be achieved, for example, by random selection of prototypes [33] or by condensing [34].

It has been repeatedly shown that randomly selected prototypes yield high quality data representations and classifiers, given sufficiently large training sets [28], [31], [33]. However, the datasets used for training of road-sign classifiers are usually rather small due to high data acquisition costs. It may thus be argued that systematic prototype selection techniques may be more beneficial for typical road-sign-classification problems than the random selection. An example of a systematic prototype selection is a validated random search. In a cross-validation fashion, a random prototype subset is drawn N times. A classifier is trained on each of the similarity representations, and its error is estimated on an independent validation set. The prototype subset that yields the minimum validation error is selected.

Apart from selection, the prototype objects may be also extracted from training instances computing, for example, the mean image per class or cluster. This technique may be beneficial in averaging out the effects of noise in individual training examples. Note that extracting a prototype object from numerous training examples does not increase the computational demands in execution.

V. EXPERIMENTS WITH TRAINABLE SIMILARITY

This section describes experiments with the proposed trainable similarity measure focusing on the training algorithm, construction of a similarity space, and the sensitivity to template displacement. Further experiments comparing the performance of various road-sign classifiers are presented in Section VI.

Contrary to studies describing monolithic road-sign classifiers covering many diverse road-sign classes in one step [5], [27], we advocate the decomposition of the road-sign-classification problem into a set of simpler subproblems based on the available prior knowledge on sign grouping [1], [10]. Such decomposition strategy allows for use of a problem-specific data representation in each of the subproblems.

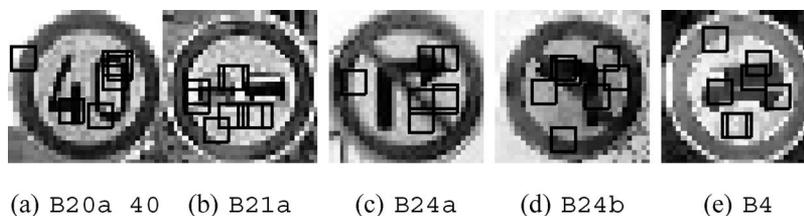


Fig. 1. Local regions derived during training of the similarity measure to five randomly selected prototypes from different classes. The regions overlay the corresponding prototype image. Road-sign classes are given under images.

Moreover, in situations where the low image resolution does not permit accurate estimation of the sign class, the hierarchical system may still provide valuable information on the rough sign category (e.g., prohibition sign).

The two experimental datasets used throughout this paper represent two situations occurring when designing hierarchies of road-sign classifiers. The dataset A contains 119 images of circular road signs from five classes (B20a 40 speed limit 40, B21a no overtaking, B24a no turning right, B24b no turning left, and B4 prohibited to trucks). This dataset represents a terminal node of a classification hierarchy with well-defined and compact classes of similar type (prohibition signs with red border and black pictograph positioned in a white center area). The dataset B corresponds to a higher level problem of separating three different types of circular road signs, namely the red–white–black prohibition signs (eight terminal sign classes), the blue–white obligatory signs (eight classes), and the red–blue signs of two classes: no stopping and no parking. This three-class dataset with 381 examples therefore illustrates a highly multimodal problem.

All images were acquired in a real environment under general illumination conditions using different digital cameras. Input color images of variable size were first converted to a gray level and rescaled to 32×32 pixel raster by a nearest neighbor interpolation.

In general, the sizes of individual local regions used within the trainable similarity measure may vary. However, in the presented set of experiments, we have fixed the region size as the external metaparameter and optimize only the region coordinates and, eventually, the region count. Apart from simplifying the search, this also allows for an effective implementation as the evaluation of local matches may be vectorized.

A. Trainable Similarity Measure

In this section, we illustrate the creation of a similarity space using the trainable similarity measure and five-class dataset A. The similarity space is spanned by dimensions measuring similarities to prototype objects. In this example, we select a single prototype image randomly from each class (see Fig. 1). Similarities to these five prototypes therefore define a five-dimensional (5-D) similarity space. This space may be now populated by available training examples as illustrated in Fig. 2. Each of the subfigures depicts the two-dimensional (2-D) scatter plot of the 5-D similarity space. Fig. 1(a) contains the similarity space constructed using the normalized cross correlation measure. Fig. 1(b) presents the similarity space derived by the trainable similarity measure S_{mean} (6×6 regions). Note that because

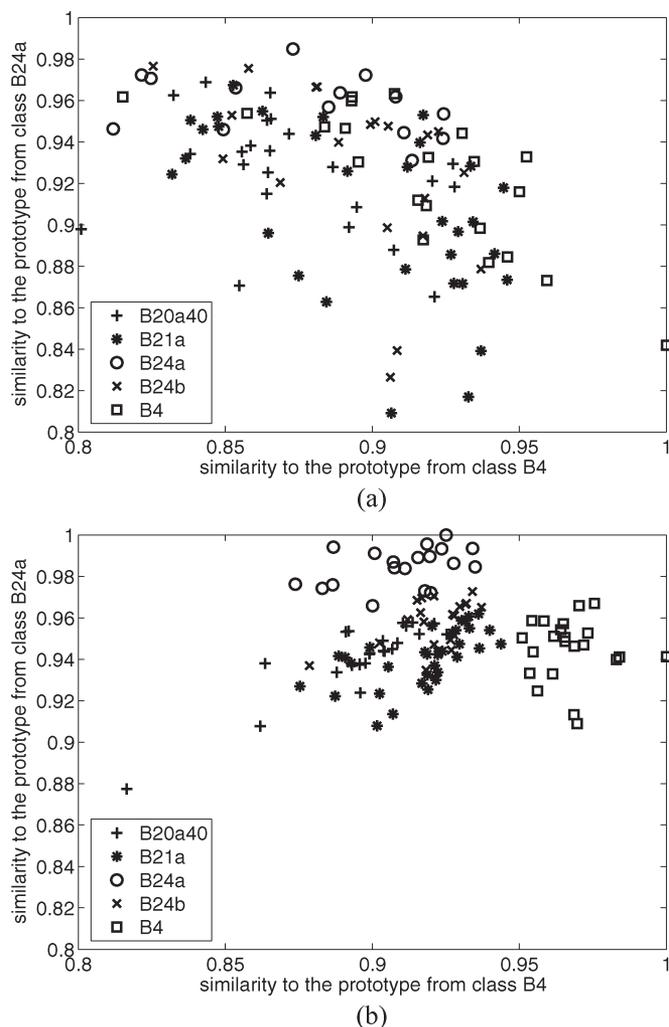


Fig. 2. (a) Scatter plots of similarity spaces using the normalized cross correlation over complete images and (b) trained measures based on local correlations using mean of local matches. Dimensions in plots measure similarity of training examples to randomly selected prototype images (see Fig. 1).

the prototype objects were selected from the training dataset, the corresponding points have unit self-similarity and may be, therefore, seen on each of the axes.

The normalized cross correlation measure results in a space with scattered observations from all five classes suffering from the presence of uninformative pixels in the road-sign images. Using the same data, the trainable similarity yields compact and well-separated classes. Each space dimension reflects only the local clues, which is important for separating the class of the corresponding prototype from other classes. Fig. 1 illustrates that trainable similarity measure, which is derived on images

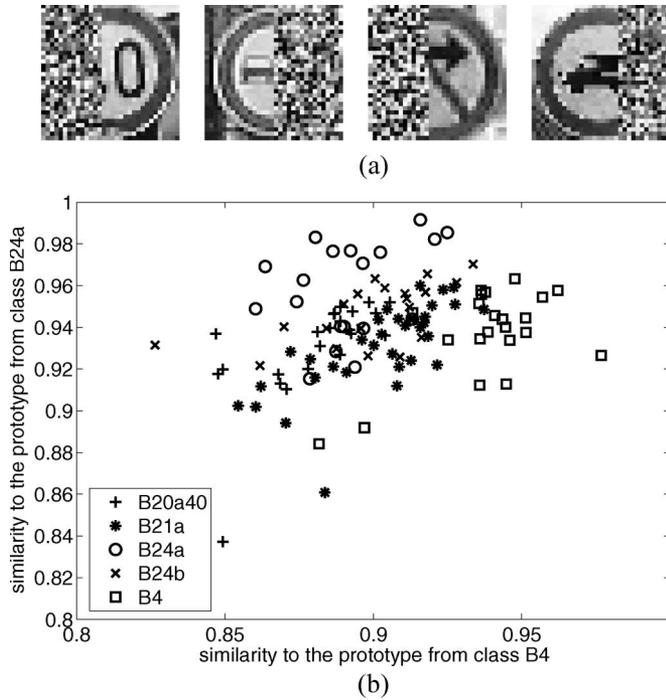


Fig. 3. Effect of occlusions on similarity space built using the trainable similarity. The points in the scatter plot represent images from dataset A, corrupted by occlusions covering 20%–50% of image by uniformly distributed random noise [examples in (a)]. The trainable similarity measure S_{mean} , derived in Section V-A with respect to the prototypes in Fig. 1, was computed for the disturbed images. (a) Examples of road sign images partially occluded by uniformly-generated random noise. (b) Trainable similarity of disturbed images to prototypes in Fig. 1.

with general BG, results in local regions that occupy only the relevant area inside the sign boards. This becomes important in cases when the road-sign detector does not provide the shape information and the image BG cannot be masked out.

B. Robustness to Partial Occlusions

Normalized cross correlation similarity is known to be very sensitive to partial object occlusions or uninformative pixels. In this section, we visualize the effect of partial occlusions on the trainable similarity. The examples in dataset A are corrupted by covering 20%–50% of image by uniformly distributed random noise. The procedure simulates occlusions by vertical objects such as lamp poles or tree trunks. The trainable similarity measures, derived in Section V-A, are now used to compute the similarity of the corrupted images to the prototypes shown in Fig. 1. Fig. 3 shows a scatter plot of a similarity space constructed identically to the one in Fig. 2(b). Note that despite significant disturbances of input images, the similarity space still retains the original class structures and thereby remains informative for the sake of classification. We conclude that due to averaging of number of local matches, the trainable similarity appears to be robust to partial occlusions.

C. Effect of Displacement on Similarity Assessment

A road-sign detector often generates partially displaced regions. Robustness to the template displacement is an impor-

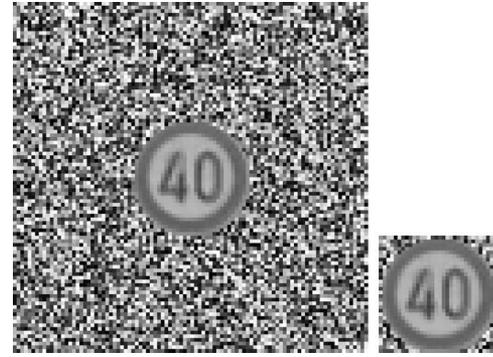


Fig. 4. Simulated patch of a traffic scene with random BG (left) and the prototype to be matched (right).

tant quality of data representation used in building a road-sign classifier. A classifier providing high accuracy in ideal conditions may become inadequate if it is sensitive to small pixel displacements of a candidate region.

In order to measure the robustness of a classifier based on the notion of image similarity, we perform the following experiment. A road-sign image is scaled and placed in the center of a bigger image filled with randomly generated pixel values obeying uniform distribution. In this way, we simulate a patch of a traffic scene with general BG. A prototype image is simulated by placing identical road-sign board with randomly generated BG pixels (see Fig. 4).

In a regular grid, the prototype image is placed over the traffic scene patch, and a similarity between the two images is computed. The resulting set of similarity values is visualized in a form of a 2-D surface in Fig. 5. The upper subfigure presents the results of the normalized cross correlation S_{Rn} and the lower subfigure the output of the trainable similarity S_{mean} . The measure was trained on the five class dataset A using a mean of the speed-limit class B20a 40 as a prototype. The forward search algorithm was used to identify the 8×8 local regions, as explained in Section III-A. Note that a set of regions derived with respect to the mean prototype [see Fig. 5(b)] is different from the regions found considering the randomly selected prototype in Fig. 1(a).

In Fig. 5(a), we can observe that normalized cross correlation produces a sharp peak centered in the best match position. The trained similarity yields a broader base with several peaks reaching even higher absolute value than the normalized cross correlation. The peak in Fig. 5(b) is asymmetric with more mass along the horizontal axis of the prototype. This may be understood by taking into account the set of regions under consideration [see the prototype in Fig. 5(b)]. When the prototype image is shifted over the scene (Fig. 4) in the horizontal direction, the local regions defined in the prototype will match highly similar scene patches. The mean of these local similarities will still attain a high value. Displacing the template in the vertical direction yields, on the other hand, less similar regions. This example illustrates, in the case of template displacement, the trainable similarity that can result in a broader peak than the normalized cross correlation.

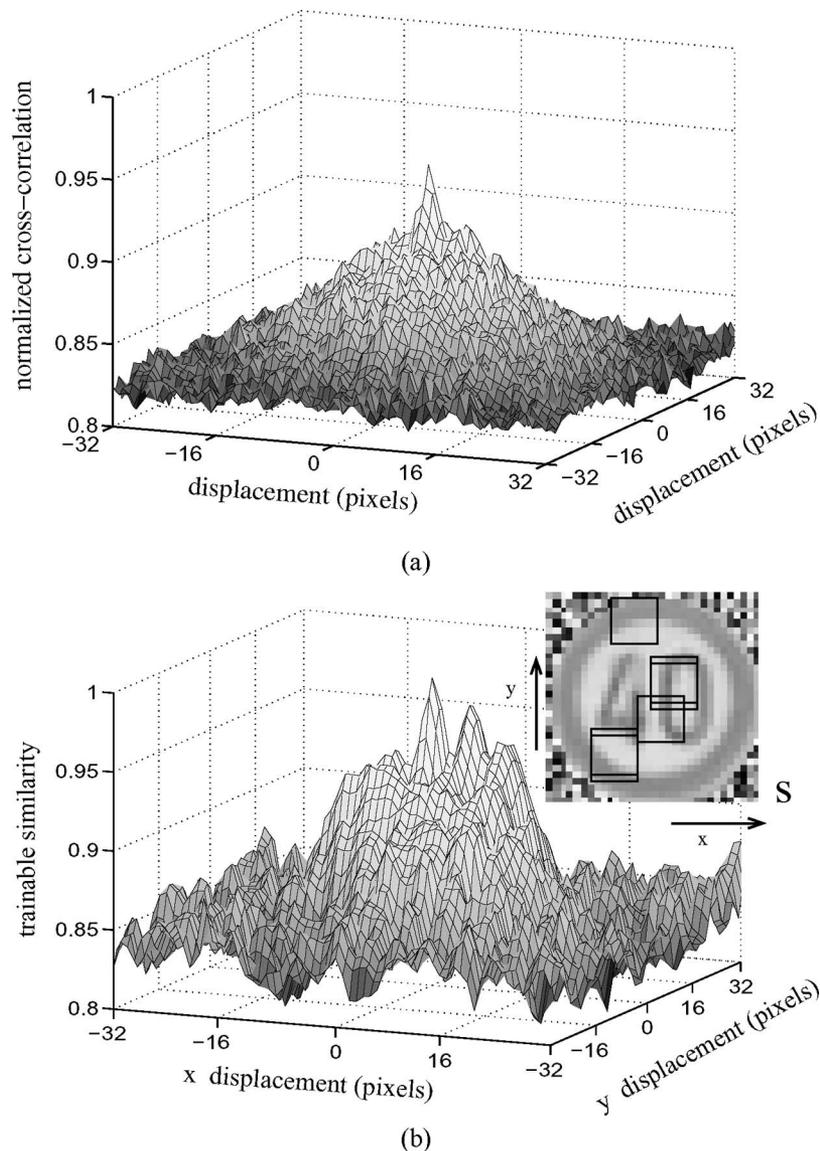


Fig. 5. Response of normalized cross correlation and trainable similarity to different template displacements. For trainable similarity, the prototype image is also shown. It is overlaid with local regions that are identified during training. (a) Normalized cross-correlation S_{Rn} . (b) Trainable similarity S_{mean} .

VI. PERFORMANCE COMPARISON OF ROAD-SIGN CLASSIFIERS

In this section, we investigate the performance of road-sign classifiers utilizing the proposed trainable similarity measure and the normalized cross correlation. For the sake of comparison, we also include two feature-based data representations. Existing studies on road-sign classification usually focus on the issue of multiclass classification accuracy and speed of prototype implementation [5], [6]. In this paper, we additionally discuss the ability of a classifier to reject nonsign examples and estimate its computational complexity in execution, which is implementation independent.

In order to evaluate the performance of different road-sign-classification strategies given a limited amount of labeled examples, a ten-fold cross-validation method is adopted. In each fold, the dataset is split into ten parts; nine of them are used for training and the remaining independent parts for the evaluation. All the steps required when building a road-sign classifier, i.e.,

representation building, feature selection, prototype selection, or training of the similarity measure are performed on the training set only. Test sets used for algorithm evaluation contain examples unseen during training. The preprocessing of the variable-sized input images is fixed for all the investigated algorithms and is, therefore, excluded from cross validation.

A. Preprocessing

As mentioned in Section V, the input variable-sized color images are preprocessed by conversion to a gray-level representation and rescaling to a fixed 32×32 pixel raster. In the set of experiments with road-sign classifiers, we consider an additional preprocessing technique: a BG removal. The road-sign detector may, depending on the used algorithm, provide information on the sign board shape. If available, the board shape may be used to mask out the sign BG as shown in Fig. 6. In order to compare algorithm performance with and without BG masking, we use the both preprocessing variants.



Fig. 6. Preprocessing by BG masking. The left-most image shows the candidate region scaled to 32×32 raster. The circular mask is given in the middle image and the final processed image in the right-most pane.

B. Data Representation

When building a similarity-based data representation, a set of prototype objects must first be selected. We consider the following prototype selection strategies, introduced in Section IV.

- Class prototype:
 - 1) Randomly selected.
 - 2) Randomly selected with 20-fold validation: A set of class prototypes is randomly selected on a 80% subset of training data and its performance is validated on the rest of the training set selecting the best set. This procedure is repeated 20 times.
 - 3) Mean prototypes.
- Mode prototypes derived by clustering: Each class is clustered using an algorithm proposed in [23]. The algorithm identifies a set of clustering solutions using the mode-seeking technique [35]. The final clustering is determined as a solution with p -percentile of a distribution of number of clusters. Because the number of clusters is not fixed by the user but estimated from the data, this method is appropriate for evaluation by cross validation. In this paper, the parameter p is set to $p = 0.95$, resulting in a large number of diverse prototypes (cluster modes).
- Full set of objects.
- Randomly selected prototypes.

Given a set of prototype objects, a similarity representation is built using either the normalized cross correlation similarity measure S_r or the trainable similarity measure S_{mean} . The trainable similarity measure derives a set of local regions using the sequential forward search algorithm.

For the sake of comparison, two feature-based data representations are also included: The first is comparable to feature sets used in our previous studies [1], [10]. In total, 41 features are considered, which have been computed from preprocessed regions: Seven Hu's moment invariants, shape characteristics R_a and θ [36], and subsampled horizontal and vertical projections of the road-sign board. A classifier is trained on a smaller feature subset refined by a sequential forward feature selection using the apparent error of the nearest neighbor classifier (1-NN) as a criterion.

The second feature-based data representation considers the raw pixel intensities directly as features in a similar way to [15] or [17]. The excessively large input dimensionality (32×32 pixel region yields a 1024-D feature space) is reduced by a principal component analysis (PCA). The PCA feature extraction was performed on a pooled class covariance matrix, retaining a fixed amount of 95% of total variance.

C. Classification

Apart from the 1-NN directly applied to the similarity-based data representation, two conventional classifiers are applied in similarity spaces: the Fisher linear discriminant (FLD) and the soft independent modeling of class analogy (SIMCA) classifier. The FLD classifier projects the input data to a lower dimensional linear subspace, maximizing the Fisher separability criterion [2], [3]. A linear discriminant is applied to the projected data. In this paper, we use a multiclass FLD classifier from [37].

The SIMCA classifier developed in the chemometric community defines a separate linear subspace for each of the classes [38] using the PCA.² A distance of an incoming observation to each class model is computed. An observation is assigned to the class with the minimum distance. The distance to a class consists of two components: a Euclidean distance to the class subspace and a Mahalanobis distance of the projected feature vector to the class mean. Both distances are normalized by the critical values of respective in-model and out-of-model data distributions. The fraction of variance preserved by the PCA was in all experiments set to 0.95.

1) *Multiclass Classification Performance:* Multiclass classification performance estimated for a set of methods using the dataset A is provided in Table I. For each method, two experiments were performed: one using only foreground (FG) pixels (with masked out BG) and the second on complete regions. The table is split in three sections, corresponding to the three investigated types of data representations. The results are given as estimated mean errors ($\hat{\mu}$) and standard deviations of the mean ($\hat{\sigma}$). From the results, summarized in Table I, we conclude the following.

- 1) When using normalized cross correlation and class prototypes, the prototype selection used is important. The first three table rows demonstrate that when using a single prototype per class, its selection significantly influences the result. Systematic selection is better than random. The results using mean class prototypes are comparable to the full nearest neighbor rule (using on average 107 prototypes per cross-validation fold).
- 2) Classifiers built in a similarity space using normalized cross correlation require tens of prototypes in order to reach low error rates. FLD built in the 30-dimensional (30-D) similarity space (randomly selected prototypes) reaches the performance comparable with the full nearest neighbor rule.
- 3) Classifiers using trainable similarity consistently reach low error rates. Note that even direct ranking of the trainable similarity to randomly selected class prototypes yields lower error than the full nearest neighbor rule using the normalized cross correlation similarity.
- 4) Different sizes of local regions do not dramatically alter the performance of classifiers based on the trainable similarity.
- 5) Trainable similarity-based classifiers are resilient to uninformative pixels. All other techniques, with the exception of the PCA-based feature extraction directly applied to

²Similar approach was adopted also by Sung and Poggio for the sake of face detection in [39].

TABLE I
SUMMARY OF CLASSIFICATION ERRORS OBTAINED BY CROSS VALIDATION ON THE FIVE-CLASS DATASET A. FOR SIMILARITY-BASED REPRESENTATIONS, THE NUMBER OF PROTOTYPES USED ON AVERAGE PER CROSS-VALIDATION FOLD IS GIVEN IN PARENTHESES

<i>data</i> <i>represent.</i>		<i>classifier</i>		<i>only foreground</i>		<i>with background</i>	
				$\bar{\mu}$ [%]	$\bar{\sigma}$ [%]	$\bar{\mu}$ [%]	$\bar{\sigma}$ [%]
relative (normalized cross- correlation)	class proto.(random), 1-NN (5)	20.4	4.6	50.1	4.9		
	class proto.(random, 20-fold val.), 1-NN (5)	11.0	2.5	39.0	4.6		
	class proto.(mean), 1-NN (5)	2.8	1.4	19.6	2.7		
	full 1-NN, (107)	2.8	2.0	11.5	3.3		
	random proto., 10-D sim.space, FLD (10)	10.7	2.3	21.5	2.7		
	random proto., 20-D sim.space, FLD (20)	7.5	1.9	12.4	2.5		
	random proto., 30-D sim.space, FLD (30)	4.4	1.5	7.1	1.8		
relative (trainable similarity)	6 × 6 pix., class proto.(random), 1-NN (5)	1.9	1.2	2.7	1.4		
	6 × 6 pix., class proto.(random), FLD (5)	1.9	1.2	2.7	1.4		
	6 × 6 pix., class proto.(random), SIMCA (5)	3.5	2.0	3.4	1.4		
	6 × 6 pix., class proto.(mean), SIMCA (5)	1.0	1.0	1.9	1.3		
	4 × 4 pix., class proto.(mean), FLD (5)	2.6	1.4	2.8	2.0		
	6 × 6 pix., class proto.(mean), FLD (5)	1.0	1.0	1.9	1.3		
	8 × 8 pix., class proto.(mean), FLD (5)	1.0	1.0	1.0	1.0		
	10 × 10 pix., class proto.(mean), FLD (5)	1.9	1.3	1.9	1.3		
absolute (feature- based)	41 features, feat.sel., FLD	10.8	4.1	16.0	3.1		
	41 features, feat.sel., SIMCA	11.2	3.1	46.9	2.5		
	41 features, feat.sel., Parzen	3.6	1.5	46.9	3.2		
	raw pixel values, PCA, FLD	0.9	0.9	2.8	1.4		

TABLE II
SUMMARY OF CLASSIFICATION ERRORS OBTAINED BY CROSS VALIDATION ON THE MULTIMODAL DATASET B WITH THREE CLASSES. FOR SIMILARITY-BASED REPRESENTATIONS, THE NUMBER OF PROTOTYPES USED ON AVERAGE PER CROSS-VALIDATION FOLD IS GIVEN IN PARENTHESES

<i>data</i> <i>representation</i>		<i>classifier</i>		<i>only foreground</i>		<i>with background</i>	
				$\bar{\mu}$ [%]	$\bar{\sigma}$ [%]	$\bar{\mu}$ [%]	$\bar{\sigma}$ [%]
relative (normalized cross- correlation)	class proto.(random, 20-fold val.), 1-NN (3)	31.4	3.1	36.3	2.3		
	class proto.(mean), 1-NN (3)	20.7	1.1	29.6	2.8		
	full 1-NN (275)	12.1	1.8	16.6	2.1		
	random proto., 10-D sim.space, FLD (10)	28.7	2.8	26.8	3.1		
	random proto., 30-D sim.space, FLD (30)	16.5	1.6	16.5	2.1		
	random proto., 50-D sim.space, FLD (50)	12.3	1.2	14.2	1.9		
	random proto., 50-D sim.space, SIMCA (50)	11.0	2.4	16.6	2.1		
	random proto., 100-D sim.space, FLD (100)	10.3	1.1	13.9	1.5		
	clustering, mode proto., FLD (26)	15.7	1.8	17.4	2.2		
	clustering, mode proto., SIMCA (26)	9.7	2.6	17.6	1.9		
relative (trainable similarity)	6 × 6 pix., class proto.(random), 1-NN (3)	22.6	2.8	20.7	2.1		
	6 × 6 pix., class proto.(mean), 1-NN (3)	14.9	2.0	26.0	2.4		
	6 × 6 pix., clust., mode proto., SIMCA (26)	11.0	1.5	13.9	2.2		
	6 × 6 pix., clust., mode proto., FLD (26)	15.2	2.2	15.8	2.1		
	8 × 8 pix., clust., mode proto., SIMCA (26)	9.7	1.6	12.8	2.1		
	8 × 8 pix., clust., mode proto., FLD (26)	15.2	2.2	15.2	2.1		
feature- based	41 features, feat.sel., FLD	20.7	3.2	27.6	3.2		
	41 features, feat.sel., Parzen	9.4	1.5	53.3	0.3		
	raw pixel values, PCA, FLD	14.0	0.9	15.0	1.7		

raw pixel intensities, are highly sensitive to general BG in images.

- 6) The feature-based representations also yield high-accuracy multiclass classification. The Parzen classifier on selected subset of features and FLD on features extracted from raw pixel intensities offer high accuracies.

Table II summarizes the multiclass classification results using the three-class multimodal dataset B.

- 1) Due to multimodal nature of the data, a single prototype per class is not sufficient for class separation. Neither prototypes selected by the validated random search nor the mean class prototypes provide a good quality data representation. However, both prototype selection techniques yield better results with the trainable similarity measure than using the normalized cross correlation.
- 2) For normalized cross correlation, using all available training examples as prototypes significantly improves performance compared to 1-NN using mean class prototypes.

- 3) Full 1-NN is outperformed by classifiers using similarity space. For normalized cross correlation, more than 50 prototypes are needed.
- 4) Mode prototypes, which are identified by clustering, yield better representations than randomly selected prototypes. On normalized cross correlation, the SIMCA classifier performs significantly better than the FLD.
- 5) Simple feature representation computed on sign FG pixels with the Parzen classifier yields the lowest classification error of 9.4%. However, when computed over the complete image, the Parzen performance deteriorates significantly.
- 6) Both relative representations benefit from cluster mode prototypes. The trainable similarity, however, exhibits more resilience to uninformative pixels than the normalized cross correlation.
- 7) Features extracted by PCA from raw pixel intensities yield moderate result.

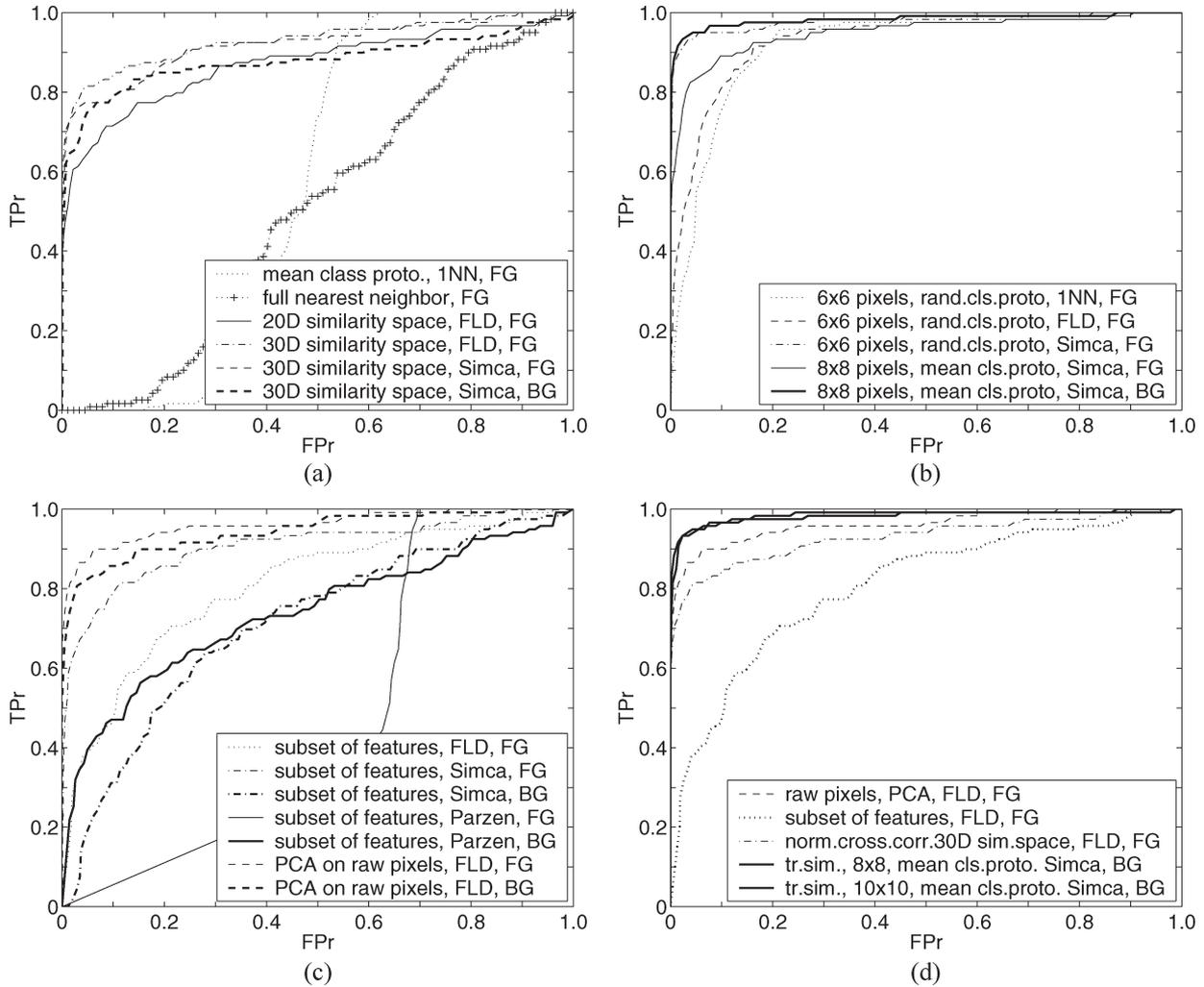


Fig. 7. ROC curves for different road-sign classifiers. Thin and thick lines denote classifiers trained on the masked sign FG and on complete images including general BG, respectively. (a) Representations based on normalized cross-correlation. (b) Representations based on trainable similarity. (c) Feature-based representations. (d) The best algorithms for each type of representation.

2) *Rejection Capability of a Classifier*: In addition to high multiclass classification accuracy, a road-sign classifier must be able to effectively reject nonsign examples. These are inevitably generated as the false alarms of the road-sign detector. Here, we evaluate the rejection performance of road-sign classifiers using a receiver operator curve (ROC) expressing the relation between the true positive ratio TP_r and the false positive ratio FP_r in a sign versus nonsign two-class problem. The following classifier outputs are thresholded and used for building of ROC curves.

- 1) Parzen classifier—maximum estimated posterior;
- 2) FLD—distance to the separation hyper plane;
- 3) 1-NN—distance to the closest training example;
- 4) SIMCA—distance to the closest class model.

The ROC curves given below were constructed from the ten-fold cross-validation results using the dataset A (Section VI). For each fold, a two-class test dataset was built with the sign and nonsign examples. While the sign examples originate from the test set of a given fold, nonsigns are formed by a separate dataset with 888 images. In order to be as close as possible to

real conditions, we use a set of nonsign regions, which are identified in a database of urban traffic scenes by the road-sign detector and designed by Líbal [8]. The detector is based on the hierarchical spatial feature matching (HSFM) algorithm utilizing the local edge information. The detector was trained on a different set of images to the data used in this paper.

Fig. 7 presents ROC curves of different road-sign classifiers. Thin lines denote classifiers trained on presegmented sign boards (i.e., FG). Thick lines represent classifiers trained on the full candidate regions with general BG.

Fig. 7(a) presents the methods, based on the normalized cross correlation. We can observe that both 1-NN using either a single-class prototype or the full set of training examples fail to reject the nonsign examples. The classifiers trained in the similarity space, however, reach similar or slightly worse rejection performance than the feature-based techniques in Fig. 7(c). It follows from experiments not shown here that the nearest neighbor rule is capable of moderate rejection performance, if the data are not normalized by mean subtraction. Its multiclass classification accuracy, however, deteriorates rapidly.

Methods based on trainable similarity measures are presented in Fig. 7(b). Although moderate rejection performance is achieved even by direct ranking of the trainable similarity (dotted line), the false positive rate remains high. The FLD, which is trained in the similarity space (dashed line), does not significantly improve the ranking results. The SIMCA classifier, on the other hand, yields very good ROC curve (dash-dotted line). Our results suggest that the high rejection performance is reached in a range of region sizes. It is also interesting that training the similarity measure using complete regions including general BG improves the rejection performance of a classifier in some cases.

Classifiers, based on feature-based representations are given in Fig. 7(c). Note that the Parzen classifier trained on features computed from FG pixels (thin solid line) exhibits very poor rejection performance although it reached low multiclass classification error of 3.6% (Table I). When trained on complete regions containing the BG (thick solid line), its rejection ability significantly improves. The classification error, however, rises to 46.9%. This result illustrates that high multiclass classification accuracy does not guarantee a good rejection performance of a classifier.

The FLD and SIMCA classifiers, which are trained using the same data representations as Parzen classifier, yield significantly better results. We conclude that these parametric models leverage the small training set better than the Parzen classifier. The best rejection performance is attained by the FLD trained in a feature space derived by PCA from raw pixel intensities. This classifier is also quite insensitive to presence of uninformative BG pixels (see Table I).

In the following, we illustrate why the data representation, which is derived by the trainable similarity measure, allows for the high rejection performance. Fig. 8 shows the 2-D scatter plot of a 5-D similarity space, constructed from test examples used in the ROC experiments above. Each dimension of the plot measures the similarity to a mean class prototype. For the sake of clarity, all the five types of signs are rendered by different markers. We can see that although the nonsign examples (marked by dots) were not used while deriving the data representation, they fall into the well-defined region with low or negative correlation with respect to the prototype images. The sign examples from the remaining three classes exhibit low correlation values to the two rendered prototypes. They are, however, analogously separated from the nonsign examples by the corresponding dimensions of the similarity space.

D. Speed of Classifier Execution

Eventually, the road-sign classifier should operate in a real-time environment, where the execution speed is of crucial importance. Measuring the execution speed is implementation and hardware dependent. We adopt an alternative strategy estimating the number of operations needed for processing of a single incoming observation. We assume an ideal implementation where all information that may be precomputed in order to speed up the algorithm execution is precomputed. Details on the estimation of the number of operations may be found in [23]. Fig. 9 depicts the generalization error in a ten-fold cross-

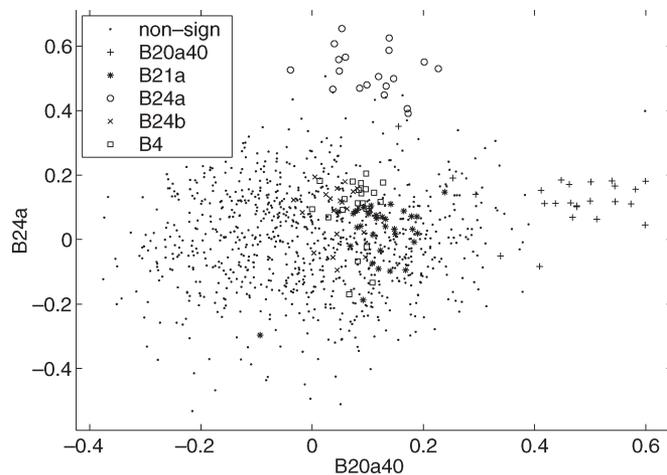


Fig. 8. Two-dimensional scatter plot of the 5-D similarity space constructed using trainable similarity to class mean prototypes. The space is filled with sign and nonsign examples unseen during training of the similarity measure. Note that nonsign examples fall into a well-defined region with low or negative correlation to prototypes and may be easily rejected.

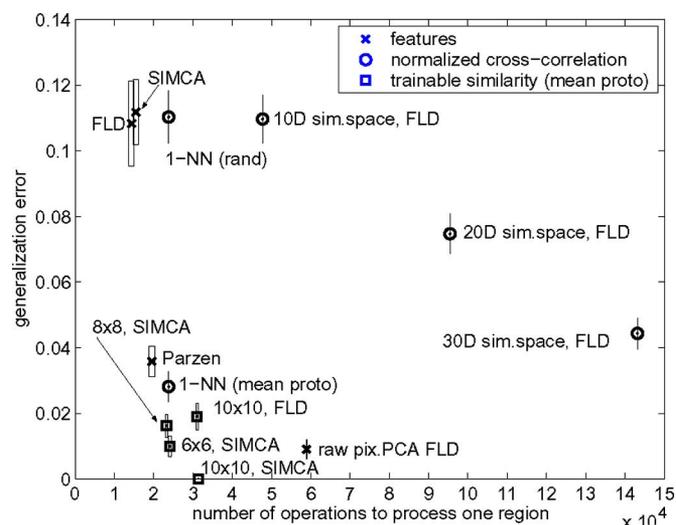


Fig. 9. Generalization error versus number of operations required for execution of a classifier on single region (including representation building). The full 1-NN classifier remains out of the plot requiring 5×10^5 operations. (Color version available online at <http://ieeexplore.ieee.org>.)

validation experiment on dataset A (Section VI) as a function of the number of operations required for processing a single candidate region. Due to BG subtraction, the algorithms operate only on 793 of the original 1024 pixels. For each method, minimum and maximum operations counts and classification errors are plotted.

The number of operations required by the classifiers based on normalized cross correlation grows linearly with the number of prototypes. The performance of the 1-NN classifier to randomly selected class prototypes may be significantly improved by training the classifiers in a similarity space. However, their computational demands become excessive due to a large number of prototypes needed to reach a moderate performance. Interestingly, the 1-NN classifier to mean class prototypes yields high accuracy, retaining the low computational demands.

Methods based on the trainable similarity reach low classification error for the fraction of computational effort of the normalized cross correlation, utilizing the same number of prototypes. The reason is that only a limited number of local image regions is employed in each comparison. Note that growing the size of local regions does not directly imply slower computation of the trainable similarity.

VII. CONCLUSION

In this paper, the novel concept of a trainable similarity was proposed as a representative building technique for road-sign classification. The proposed trainable similarity extends the commonly used normalized cross correlation approach, which is global and symmetric, by deriving a local measure that is specific to a particular prototype. This involves computing the similarity between an observed image and a labeled prototype example based on multiple matches in local image regions. The set of regions is derived by training from a set of labeled examples, leveraging the available knowledge of the class membership of prototype objects. The proposed method incorporates locality typical for feature extraction into the design of a similarity measure. However, the identified prototype-specific regions are not used directly as features but rather are employed for the sake of image matching. The resulting representation therefore exhibits desired properties of a similarity measure such as retaining low values for nonsign examples unseen during training. This is crucial for robust rejection of false alarms introduced by the road-sign detector.

Although this representation building strategy is also applicable to other image recognition tasks, it specifically suits the road-sign-classification problem for several reasons.

First, the road-sign classes are often multimodal, implying that representation of individual sign variants should reflect specific local clues. The proposed approach facilitates the design of such representations by training a specific similarity measure to each of the prototypes.

Second, some road-sign detectors may not provide information on the sign board shape, and therefore, the uninformative sign BG cannot be removed by preprocessing. Additionally, uninformative pixels may also be introduced by slight displacement of the detector template or by sign occlusions. The trainable similarity appears to be more robust in these situations than the normalized cross correlation and yields more informative data representation for the sake of road-sign classification.

Finally, as fewer prototypes are needed and only a pixel subset is used in each image comparison, the resulting road-sign classifiers also execute significantly faster than algorithms based on the normalized cross correlation.

ACKNOWLEDGMENT

The authors would like to thank T. Landgrebe and C. Lai for their fruitful comments on the paper.

REFERENCES

- [1] P. Paclík, J. Novovičová, P. Pudil, and P. Somol, "Road sign classification using laplace kernel classifier," *Pattern Recognit. Lett.*, vol. 21, no. 13/14, pp. 1165–1173, 2000.
- [2] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Hoboken, NJ: Wiley, 2001.
- [3] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 2nd ed. New York: Academic, 2003.
- [4] M. Lalonde and Y. Li, "Survey of the state of the art for sub-project 2.4: Road sign recognition," in "Collection scientifique et technique," Centre de recherche informatique de Montreal, Montreal, QC, Canada, Tech. Rep., CRIM-IT-95/09-35, 1995.
- [5] G. Piccioli, E. De Micheli, P. Parodi, and M. Campani, "Robust method for sign detection and classification," *Image Vis. Comput.*, vol. 4, pp. 209–223, 1996.
- [6] S. H. Hsu and C. L. Huang, "Road sign interpretation using matching pursuit method," *Image Vis. Comput.*, vol. 9, pp. 119–129, 2001.
- [7] A. de la Escalera, J. M. Armigol, and M. Mata, "Traffic sign recognition and analysis for intelligent vehicles," *Image Vis. Comput.*, vol. 11, no. 3, pp. 247–258, 2003.
- [8] V. Lbal, "Road signs recognition system," Ph.D. dissertation, Czech Technical University, Prague, Czech Republic, 2001.
- [9] L. Priese, J. Klieber, R. Lakmann, V. Rehrmann, and R. Schian, "New results on traffic sign recognition," in *Proc. IEEE Intell. Veh. Symp.*, Paris, France, Oct. 24–26, 1994, pp. 249–254.
- [10] P. Paclík and J. Novovičová, "Road sign classification without color information," in *Proc. 6th ASCI Conf., Lommel, Belgium*, L. van Vliet, J. Heijnsdijk, T. Kielman, and P. Knijnenburg, Eds. Delft, The Netherlands: ASCI, 2000.
- [11] P. Douville, "Real-time classification of traffic signs," *Real-Time Imaging*, vol. 6, no. 3, pp. 185–193, Jun. 2000.
- [12] R. Dahyot and P. Charbonier, "Robust visual recognition of colour images," in *Proc. CVPR*. Hilton Head Island, SC, Jun. 2000, pp. 685–690.
- [13] D. G. Shaposhnikov, G. A. V. Podladchikova, N. Lubov, N. A. Shevtsova, K. Hong, and X. Gao, "Road sign recognition by single positioning of space-variant sensor window," in *Proc. 15th Int. Conf. Vis. Interface*, Calgary, AB, Canada, May 27–29, 2002, pp. 213–217.
- [14] L. Priese, R. Lakmann, and V. Rehrmann, "Ideogram identification in a realtime traffic sign recognition system," in *Proc. IEEE Intell. Veh. Symp.*, Detroit, MI, Sep. 25–26, 1995, pp. 310–314.
- [15] A. de la Escalera, L. E. Moreno, M. A. Salichs, and J. M. Armigol, "Road traffic sign detection and classification," *IEEE Trans. Ind. Electron.*, vol. 44, no. 6, pp. 848–859, Dec. 1997.
- [16] U. Franke, D. Gavrilu, S. Gorzig, F. Lindner, F. Paetzold, and C. Wohler, "Autonomous driving goes downtown," *IEEE Intell. Syst.*, vol. 13, no. 6, pp. 40–48, Nov./Dec. 1998.
- [17] D. Gavrilu, "Traffic sign recognition revisited," in *Proc. 21st DAGM Symp. Mustererkenn.*, 1999, pp. 86–93.
- [18] M. Betke and N. C. Makris, "Fast object recognition in noisy images using simulated annealing," in *Proc. Int. Conf. Comput. Vis.*, 1996, pp. 523–530.
- [19] J. Miura, T. Kanda, and Y. Shirai, "An active vision system for real-time traffic sign recognition," in *Proc. IEEE Intell. Transp. Syst.*, Dearborn, MI, 2000, pp. 52–57.
- [20] R. C. Gonzales and R. E. Woods, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1992.
- [21] G. S. Cox, "Designing hypothesis tests for digital image matching," Ph.D. dissertation, Digital Image Signal Proc. Group, Univ. Cape Town, Cape Town, South Africa, Aug. 2000. [Online]. Available: <http://www.dip.ee.uct.ac.za/Publications/Theses>
- [22] L. Guibert, Y. Petillot, and J. L. de Bougrenet de la Tocnaye, "Real-time demonstration of an on-board nonlinear joint transform correlator system," *Opt. Eng.*, vol. 36, no. 3, pp. 820–824, Mar. 1997.
- [23] P. Paclík, "Building road sign classifiers," Ph.D. dissertation, Czech Techn. Univ., Prague, Czech Republic, 2004. [Online]. Available: <http://prsysdesign.net/pavel>
- [24] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1982.
- [25] P. Viola and M. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.
- [26] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2005, pp. 255–260.
- [27] A. de la Escalera, J. M. Armigol, J. M. Pastor, and F. J. Rodriguez, "Visual sign information extraction and identification by deformable models for intelligent vehicles," *IEEE Trans. Int. Transp. Syst.*, vol. 5, no. 2, pp. 57–68, Jun. 2004.
- [28] R. P. W. Duin, D. de Ridder, and D. M. J. Tax, "Experiments with object based discriminant functions; a featureless approach to pattern recognition," *Pattern Recognit. Lett.*, vol. 18, no. 11/13, pp. 1159–1166, 1997.

[29] C. Tzomakas, "Contributions to the visual object detection and classification," Ph.D. dissertation, Ruhr Univ., Bochum, Germany, 1999.

[30] R. P. W. Duin, E. Pekalska, and D. de Ridder, "Relational discriminant analysis," *Pattern Recognit. Lett.*, vol. 20, no. 11/13, pp. 1175–1181, 1999.

[31] E. Pekalska, R. P. W. Duin, and P. Pačlík, "Prototype selection for dissimilarity-based classifiers," *Pattern Recognit.*, vol. 39, no. 2, pp. 189–208, 2005.

[32] P. Pačlík and R. P. W. Duin, "Dissimilarity-based classification of spectra: Computational issues," *Real-Time Imaging*, vol. 9, no. 4, pp. 237–244, Aug. 2003.

[33] E. Pekalska, P. Pačlík, and R. P. W. Duin, "A generalized kernel approach to dissimilarity based classification," *J. Mach. Learn. Res.*, vol. 2, no. 2, pp. 175–211, 2002.

[34] C. Chen and A. Jóźwik, "A sample set condensation algorithm for the class sensitive artificial neural network," *Pattern Recognit. Lett.*, vol. 17, no. 8, pp. 819–823, Jul. 1996.

[35] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.

[36] W. Pratt, *Digital Image Processing*, 2nd ed. Hoboken, NJ: Wiley, 1991.

[37] R. P. W. Duin, P. Juszczak, D. de Ridder, P. Pačlík, E. Pekalska, and D. M. J. Tax, "PR-Tools 4.0, a Matlab Toolbox for Pattern Recognition," ICT Group, TU Delft, Delft, The Netherlands, Jan. 2004. Tech. Rep. [Online]. Available: <http://www.prtools.org>

[38] S. Wold and M. Sjostrom, "SIMCA: A method for analysing chemical data in terms of similarity and analogy," in *Proc. ACS Symp. Series 52*, 1977, pp. 243–282.

[39] K. K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 1, pp. 39–51, Jan. 1998.



Jana Novovičová received the M.S. degree in mathematics and the *Rerum Naturalium Doctoris* degree in mathematical statistics, both from Charles University, Prague, Czech Republic, and the Ph.D. degree in theoretical cybernetics from Institute of Information Theory and Automation, Czech Academy of Sciences (UTIA CAS), Prague, in 1982 and 1983, respectively.

She is with the Department of Pattern Recognition, UTIA CAS, and an Associate Professor of Informatics with the Faculty of Transportation Sciences, Czech Technical University, Prague. Her research interests include statistical approaches to pattern recognition (feature selection, classification, mixture models) and text document classification.



Robert P. W. Duin (M'04) received the M.Sc. and Ph.D. degrees in applied physics, focusing on the design and evaluation of systems for image processing and pattern recognition, both from Delft University of Technology, Delft, The Netherlands, in 1978.

Currently, he is leading a group of young researchers in this area as an Associate Professor of the Faculty of Electrical Engineering, Mathematics, and Computer Science of Delft University of Technology.

Dr. Duin served as an Associate Editor for various journals like the *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE* and *Pattern Recognition Letters*. He was on the program committees of numerous conferences and coorganized workshops in the field of statistical pattern recognition in his responsibility as the chairman of the Technical Committee of the International Association for Pattern Recognition (IAPR) for this field. He has coauthored more than 250 scientific publications, including eight books (three monographs and five collections) and 60 journal papers. He is a member of the IAPR.



Pavel Pačlík received the M.S. and Ph.D. degrees in transportation engineering, focusing on road-sign classification, from Czech Technical University, Prague, Czech Republic, in 1998 and 2004, respectively.

Between 1998 and 2000, he was with the Institute of Information Theory and Automation, Czech Academy of Sciences (UTIA CAS), Prague. Between 2000 and 2003, he was with the Pattern Recognition Group at the Delft University of Technology, Delft, The Netherlands. Since 2004, he has

been with the Information and Communication Theory Group. His research interests span various aspects of designing pattern recognition systems such as representation-building strategies, classifier design, clustering, multicriteria system evaluation, texture classification, and spectral imaging.

Dr. Pačlík is a member of International Association for Pattern Recognition (IAPR).