ELSEVIER

Pattern Recognition Letters 21 (2000) 1165–1173

Pattern Recognition Letters

www.elsevier.nl/locate/patrec

Road sign classification using Laplace kernel classifier ☆

P. Paclík [a,b,c,*], J. Novovičová [a,b,c], P. Pudil [b,c], P. Somol [b,c]

[a] *Faculty of Transportation Sciences, Czech Technical University, 110 00, Prague, Czech Republic*

[b] *Department of Pattern Recognition, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, 182 08, Prague 8, Czech Republic*

[c] *Joint Laboratory of the Academy of Sciences and Faculty of Management, University of Economics, Jindřichův Hradec, Czech Republic*

Abstract

Driver support systems (DSS) of intelligent vehicles will predict potentially dangerous situations in heavy traffic, help with navigation and vehicle guidance and interact with a human driver. Important information necessary for traffic situation understanding is presented by road signs. A new kernel rule has been developed for road sign classification using the Laplace probability density. Smoothing parameters of the Laplace kernel are optimized by the pseudo-likelihood cross-validation method. To maximize the pseudo-likelihood function, an Expectation–Maximization algorithm is used. The algorithm has been tested on a dataset with more than 4900 noisy images. A comparison to other classification methods is also given. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Road sign recognition; Kernel density estimation; Expectation–maximization algorithm

1. Road sign recognition

In an intelligent vehicle a driver support system (DSS) should work as a driver copilot, continuously monitoring the driver, vehicle and the environment in order to facilitate human decisions about immediate vehicle guidance and navigation (Nagel). To be able to help the driver with decision making, the DSS must understand the current traffic situation. Therefore, it should create and maintain a model of its neighborhood. Due to the dominant role of visual information for the human driver, computer vision methods are often used in intelligent vehicle prototypes for the creation of such model. Road signs offer, among the other traffic devices, a lot of important information about the current traffic situation. Two basic road sign groups exist – ideogram-based and text-based signs. While the first group uses simple ideographs to express the sign meaning, the second one contains road signs with texts, arrows and other symbols. This article is concerned with the recognition of ideogram-based road signs using statistical pattern recognition methods. A comprehensive study of road sign recognition presented by Lalonde and Li (1995) compiles information about related algorithms, research groups and results. Several research projects dealing with the road sign recognition have been reported. A few of them have led to intelligent vehicle prototypes

☆ The final work on the paper has been done in the Pattern Recognition Group, Delft University of Technology, P.O. Box 5046, 2600 GA, Delft, Netherlands.

* Corresponding author. Present address: Pattern Recognition Group, Delft University of Technology, P.O. Box 5046, 2600, GA Delft, Netherlands.

E-mail address: pavel@ph.tn.udelft.nl (P. Paclík).

0167-8655/00/$ - see front matter
PII: S0167-8655(00)00078-7

(e.g. VITA II vehicle developed by the research team at the University Koblenz–Landau together with *Daimler–Chrysler*, Priese et al., 1994). An often used approach for road sign recognition is a correlation method. A normalized image is created for each road sign type. It is applied as a template to a number of places in the traffic scene image. Template positions with the highest similarity values are then labeled by the corresponding sign codes. As the correlation method combines both the detection and classification stages it is an efficient procedure for the fast recognition of a few sign types. On the other hand, a general method separating the sign detection and classification steps may be more convenient as the number of sign types grows. The most common approach to road sign detection is based on a color segmentation method (Priese et al., 1994). The classification is then performed by a neural network. The detected image region is used as network input and image pixels are taken directly as features (Escalera et al., 1997; Franke et al., 1998).

There are some issues specific to the recognition of road signs:

- The recognition of objects in outdoor scenes is difficult due to variable illumination conditions.
- Images acquired from a moving car suffer from car vibrations and motion blur.
- Sign boards are often deteriorated by weather conditions, scratches and dust.
- There exist international standards, but real road signs considerably differ from them (see Fig. 1). The road sign classifier must take into account many sign variants. It is necessary to provide a large set of real training samples – standards themselves are not a sufficient source for classifier learning.
- The recognition method must be effective enough to be implemented in a real-time environment.
- No standard databases of road signs for evaluation of particular classification method exist (most of the research is commercial and there is no access to such resources).

This paper describes the classification module of the road sign recognition system (RS^2) which has been designed at the Faculty of Transportation Sciences, CTU Prague. Contrary to most of the presented studies, RS^2 uses local orientations of edges in the image for the road sign detection (Líbal et al., 1996, 1997, 1998). The detection algorithm searches the traffic scene image for geometrical shapes corresponding to road sign boards. The search is performed by a hierarchical template matching procedure. The detection template is able to find geometrical shapes rotated in $\pm 5°$ range from the basic position. The size of detected objects in the traffic scene image changes from 15 to 150 pixels. Road sign boards may be also partially occluded (missing triangle corner or part of a circle border does not influence the detection result). However, the algorithm does not respond to strong shape distortions at all.

Fig. 1. Differences between European road signs (sign A12 "*Children*").

2. Classification algorithm

The input of the RS^2 classification module is a list of candidate regions containing some structures resembling the road sign boards. The goal is to label these regions by the appropriate road sign codes or to reject them. The coarse meaning of the road sign (e.g. warning or prohibition) is presented by the sign shape and the color combination. The exact sign meaning is then specified by the ideogram itself. This a priori knowledge is used for the decomposition of the whole recognition problem into several smaller ones. Therefore, the classification module of RS^2 works as a

decision tree with several node classifiers (Paclík, 1998). The decision tree approach has several advantages to the single-classifier method. The first one is the reduction of the class count per decision tree node. Moreover, each particular classifier may exploit the most descriptive features for its task. Satisfactory classification results are also reached using a smaller number of features compared to a single-classifier (Paclík and Novovičová, 2000). The misclassification risk between different sign groups is reduced as the decision is made by a multi-stage system. This is a valuable property as the exchange of e.g. the *closed to all vehicles* sign with *no parking* is a fatal system error. An important feature of the decision tree approach is also the existence of *partial results*. Small images of more distant signs often lack clear ideogram data. The decision tree then reports at least the road sign type (e.g. prohibition). The rejection of many false alarms is also made at early tree levels.

2.1. Color segmentation

The color segmentation method is used to move from the input RGB color space to task specific colors. There exist advanced segmentation methods which are robust but also have considerable computational demands (Priese et al., 1994). A compromise between segmentation reliability (robustness) and speed has to be made. The hue saturation value (HSV) color space is used because of its similarity to the human perception of colors. It is a desirable feature as the segmentation algorithm separates six basic colors used in the road sign design (white, black, red, blue, green and yellow). To segment achromatic colors (white and black) the value component of HSV color model is thresholded. Other colors are obtained by thresholding the hue component (Aldon and Pujas, 1995). Thresholds were setup using a set of real traffic scenes with variable illumination conditions. The segmentation algorithm is, in fact, pixel-based classification into six classes. By this method, even adversely illuminated road sign boards are processed correctly and the algorithm is very fast. However, wrong color segmentation has fatal consequences to the classification result.

2.2. Feature vector construction

Features for the statistical pattern classifier are computed on binary images of the road sign interior. Colors to be binarized depend on the particular road sign group (e.g. white for obligatory signs or black for warning signs). Images on the classifier input may be rotated in a pre-defined range $\pm 5^\circ$ given by the detection template. On the other hand, the input image size varies considerably and used features must be therefore invariant to the scale change. Several moment invariant features have been used. The *unscaled spatial moment* of the order m, n ($F(j,k)$ is a binary image function) is:

$$M(m,n) = \sum_{j=1}^{J} \sum_{k=1}^{K} (x_k)^m (y_j)^n F(j,k). \tag{1}$$

The translation-invariant *unscaled central moment* of the order m, n is calculated using expression:

$$U_U(m,n) = \sum_{j=1}^{J} \sum_{k=1}^{K} [x_k - \bar{x}_k]^m [y_j - \bar{y}_j]^n F(j,k), \tag{2}$$

where $\bar{x}_k$ and $\bar{y}_j$ are image centroid coordinates. The scale change invariant *normalized unscaled central moments* V has been used which is given by the formula:

$$V(m,n) = \frac{U_U(m,n)}{[M(0,0)]^\alpha}, \quad \text{where } \alpha = \frac{m+n}{2} + 1, \tag{3}$$

where $M(0,0)$ stands for the image size. Another feature useful especially for the separation of circular objects is *compactness*. It is calculated using binary object area A_o and perimeter P_o in the following way:

$$\text{comp} = \frac{P_o^2}{4\pi A_o}. \tag{4}$$

For circles, compactness comes near unity while for oblong objects it takes value $\text{comp} \in (1.0, \infty)$. The perimeter is approximated by the pixel count of the object boundary which is constructed by the methods of mathematical morphology.

2.3. Laplace kernel classifier

Let us define the classification problem as an allocation of the feature vector $\boldsymbol{x} \in \mathbb{R}^D$ to one of C mutually exclusive classes knowing that the class of $\boldsymbol{x}$, denoted by ω, takes values in $\Omega = \{\omega_1, \ldots, \omega_C\}$ with probabilities $P(\omega_1), \ldots, P(\omega_C)$, respectively and that $\boldsymbol{x}$ is a realization of a random vector $\boldsymbol{X}$ characterized by a conditional probability density function $f(\boldsymbol{x}\,|\,\omega)$, $\omega \in \Omega$.

With the usual kernel approach to classification (Devroye et al., 1996; Sain, 1994), the unknown class conditional densities in the Bayes rule are replaced by the kernel density estimates obtained from the independent training data $\boldsymbol{x}_1^\omega, \ldots, \boldsymbol{x}_{N_\omega}^\omega$, $\omega \in \Omega$. The associated sample-based decision rule is therefore a plug-in version of the Bayes rule with the kernel density estimates used in the place of the class conditional densities. A nonparametric estimate of the ωth class conditional density $f(\boldsymbol{x}\,|\,\omega)$ provided by the kernel method is

$$\hat{f}(\boldsymbol{x}\,|\,\omega) = \frac{1}{N_\omega h_\omega^D} \sum_{i=1}^{N_\omega} K\left(\frac{\boldsymbol{x} - \boldsymbol{x}_i^\omega}{h_\omega}\right), \tag{5}$$

where $K(\cdot)$ is a kernel function that integrates to one and h_ω is a smoothing parameter (Devroye et al., 1996). In most applications, the kernel K is fixed and the smoothing parameter h_ω is a function of the ωth training set of the size N_ω, such that $\lim_{N_\omega \to \infty} h_\omega(N_\omega) = 0$ and $\lim_{N_\omega \to \infty} N_\omega h_\omega(N_\omega) = \infty$. Usually, the kernel $K(\cdot)$ is required to be nonnegative and symmetric. If $K(\boldsymbol{x}) \geqslant 0$ then the kernel density estimate $\hat{f}(\boldsymbol{x}\,|\,\omega)$ can be interpreted as a mixture of N_ω component densities in equal proportions. Let us consider the following multivariate product kernel estimate of $f(\boldsymbol{x}\,|\,\omega)$

$$\hat{f}(\boldsymbol{x}\,|\,\omega) = \frac{1}{N_\omega h_{\omega 1} \ldots h_{\omega D}} \sum_{i=1}^{N_\omega} \left\{ \prod_{j=1}^{D} K\left(\frac{x_j - x_{ij}^\omega}{h_{\omega j}}\right) \right\}$$

where x_j is the jth component of the vector $\boldsymbol{x}$ and $\boldsymbol{x}_i^\omega = (x_{i1}^\omega, \ldots, x_{iD}^\omega)$, $i = 1, \ldots, N_\omega$. It means that the same univariate kernel K is used in each dimension but with a different smoothing parameter $h_{\omega j}$ for each dimension. The choice for the univariate kernel function investigated here is the *Laplace* density function

$$f_L(x;\ \mu, \sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|x - \mu|}{\sigma} \right), \tag{6}$$

where $x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma \in (0, \infty)$.

Therefore, the Laplace kernel estimate of $f(\boldsymbol{x}\,|\,\omega)$ becomes

$$\hat{f}(\boldsymbol{x}\,|\,\omega) = \frac{1}{N_\omega} \sum_{i=1}^{N_\omega} \prod_{j=1}^{D} \frac{1}{2h_{\omega j}} \exp\left(-\frac{|x_j - x_{ij}^\omega|}{h_{\omega j}} \right). \tag{7}$$

We can rewrite Eq. (7) in the form

$$\hat{f}(\boldsymbol{x}\,|\,\omega) = \frac{1}{N_\omega} \sum_{i=1}^{N_\omega} f_{Li}\left(\boldsymbol{x}_i;\ \boldsymbol{x}_i^\omega, \boldsymbol{H}_\omega\right), \tag{8}$$

where $\boldsymbol{H}_\omega$ is $D \times D$ diagonal matrix with diagonal elements $h_{\omega 1}, \ldots, h_{\omega D}$ respectively, common to all densities f_{Li}, $i = 1, \ldots, N_\omega$.

2.4. Estimation of smoothing parameters

As the choice of the kernel function is not so important, the usual approach in constructing $\hat{f}(\boldsymbol{x}\,|\,\omega)$ is to fix the kernel K in Eq. (5), and then asses the smoothing parameters from the observed data (e.g. McLachlan, 1992). Appropriate selection of the smoothing parameters is crucial in the estimation process. The dependence of the kernel estimator performance on the smoothing parameters has led to many proposals (for example mean squared error or integrated square bias criteria). The standard approach for the determination of unknown parameters $h_{\omega 1}, \ldots, h_{\omega D}$ in the kernel estimate (Eq. (8)), postulated for the ωth class conditional density from given data $\boldsymbol{x}_1^\omega, \ldots, \boldsymbol{x}_{N_\omega}^\omega$, is to use maximum likelihood (ML) estimation. To compute the ML estimates of the unknown parameters we maximize the corresponding log-likelihood function.

$$L = \sum_{k=1}^{N_\omega} \ln \hat{f}(\boldsymbol{x}_k^\omega\,|\,\omega). \tag{9}$$

The log-likelihood function L of the kernel estimate given in Eq. (8) with the smoothing matrix $\boldsymbol{H}_\omega$ is known to attain an infinite maximum for $|\boldsymbol{H}_\omega| \to 0$, because $\hat{f}(\boldsymbol{x}\,|\,\omega)$ approaches zero at all

$\boldsymbol{x}$ except at $\boldsymbol{x} = \boldsymbol{x}_j^\omega,\ j = 1, 2, \ldots, N_\omega$, where it is $1/N_\omega$ times the Dirac delta function. This undesirable property can be removed by using the cross-validated log-likelihood (Duin, 1976)

$$L(\boldsymbol{H}_\omega) = \sum_{k=1}^{N_\omega} \ln \hat{f}_{-k}(\boldsymbol{x}_k^\omega \,|\, \omega), \tag{10}$$

where

$$\hat{f}_{-k}(\boldsymbol{x}_k^\omega \,|\, \omega) = \frac{1}{N_\omega - 1} \sum_{\substack{i=1 \\ i \neq k}}^{N_\omega} f_L(\boldsymbol{x}_k^\omega;\ \boldsymbol{x}_i^\omega, \boldsymbol{H}_\omega) \tag{11}$$

denotes the kernel density estimate $\hat{f}(\boldsymbol{x} \,|\, \omega)$, formed from $\boldsymbol{x}_i^\omega, i = 1, 2, \ldots, N_\omega, i \neq k$. In order to maximize the criterion in Eq. (10), we can modify the Expectation–Maximization (EM) algorithm (Dempster et al., 1977) as follows:

E-step:

$$p^{(t)}(\boldsymbol{x}_i^\omega \,|\, \boldsymbol{x}_k^\omega) = \frac{f_L(\boldsymbol{x}_k^\omega;\ \boldsymbol{x}_i^\omega, \boldsymbol{H}_\omega)}{\sum_{\substack{i=1 \\ i \neq k}}^{N_\omega} f_L(\boldsymbol{x}_k^\omega;\ \boldsymbol{x}_i^\omega, \boldsymbol{H}_\omega)} \tag{12}$$

M-step:

$$h_{\omega l}^{(t+1)} = \frac{1}{N_\omega} \sum_{k=1}^{N_\omega} \sum_{\substack{i=1 \\ i \neq k}}^{N_\omega} p^{(t)}(\boldsymbol{x}_i^\omega \,|\, \boldsymbol{x}_k^\omega) |x_{kl}^\omega - x_{il}^\omega|, \tag{13}$$

where $t = 0, 1, \ldots$.

3. Algorithm implementation

The classification algorithm with the Laplace kernel rule is presented as algorithm 1. In order to estimate the density function faster, the Eq. (7) has been rewritten as

$$\hat{f}(\boldsymbol{x} \mid \omega) = \frac{1}{2^D N_\omega \prod_{k=1}^{D} h_{\omega k}} \sum_{i=1}^{N_\omega} \exp\left(- \sum_{j=1}^{D} \frac{| x_j - x_{ij}^\omega |}{h_{\omega j}} \right)$$

The operator "./" on line 10 denotes division of corresponding vector elements. The EM algorithm for estimation of the smoothing parameters by maximization of the cross-validated log-likelihood function is given as algorithm 2.

Algorithm 1: Kernel classifier with Laplace kernel

```
1:  input: vector x (unknown pattern)
2:  output: class code
3:  training set: patterns x_1^{ω_c}, ..., x_{N_{ω_c}}^{ω_c} for
    classes ω_c, c = 1, ..., C
4:  parameters: smoothing vector h; rejection threshold reject
5:  max = 0; maxclass = nil
6:  for all classes ω_c ∈ training set T do
7:      classcontrib = 0
8:      for all patterns x_i^{ω_c},
        i = 1, ..., N_{ω_c} do
9:          work = abs(x − x_i^{ω_c})
10:         work = work ./ h_c
11:         classcontrib +
              = exp(− Σ_{j=1}^{D} work_j)
12:     end for
13:     classcontrib / = 2^D · N_{ω_c} · Π_{j=1}^{D} h_cj
14:     if classcontrib > reject and
        classcontrib > max then
15:         max = classcontrib
16:         maxclass = ω_c
17:     end if
18: end for
19: return: maxclass
```

Considerable acceleration of classification has been reached using *the sample rejection method*. The method assumes that wrong decisions are characterized by high values of the sum inside the exponential (line 11, algorithm 1) while proper decisions are characterized by lower ones. Therefore, if the sum exceeds some threshold s_r for a particular pattern $\boldsymbol{x}_i^{\omega_c}$, the pattern is rejected from further processing as being too distant. The modification of pattern loop is presented as algorithm 3.

The rejection threshold s_r is set up for particular dataset according to the analysis of the histogram of sum values for *proper* and *wrong* classifier decisions. Although both groups overlap for real data, a value of s_r separating certainly void decisions from good ones may be found. It follows from experiments that the classification may be speeded up for about 20% by the proper s_r setting without any impact to the classification results.

Algorithm 2: EM algorithm for smoothing parameters optimization

input: classes $\omega_c, c = 1, \ldots, C$; D-dimensional patterns $\boldsymbol{x}_1^{\omega_c}, \ldots, \boldsymbol{x}_{N_{\omega_c}}^{\omega_c}$ for every class ω_c;
output: smoothing vectors $\boldsymbol{h}_c$, $c = 1, \ldots, C$
parameters: maximum difference between two following estimates *dif*
$h_{cd} = 1.0, c = 1, \ldots, C, d = 1, \ldots, D$ // *init*.
for all classes ω_c such that $c = 1, \ldots, C$ **do**
  $last_d = 100.0$, for $d = 1, \ldots, D$
  repeat
    // *fill density matrix f*
    for all patterns $\boldsymbol{x}_i$ and $\boldsymbol{x}_k$ such that $i, k = 1, \ldots, N_{\omega_c}, i \neq k$ **do**
      work $= \mathrm{abs}(\boldsymbol{x}_i - \boldsymbol{x}_k)$
      work = **work** $./\ \boldsymbol{h}_c$
      $f(i,k) = \exp(-\sum_{j=1}^{D} work_j)$
      $f(i,k) = f(i,k)/(2\prod_{j=1}^{D} h_{cj})$
    end for
    // *combine E- and M-steps*
    for all features $d,\ d = 1, \ldots, D$ **do**
      $temp = 0$
      for all patterns $\boldsymbol{x}_i$ and $\boldsymbol{x}_k$ such that $i, j = 1, \ldots, N_{\omega_c}, i \neq j$ **do**
        $p = f(i,k)/\sum_{m=1, m\neq i}^{N_{\omega_c}} f(i,m)$
        $temp+ = \mathrm{abs}(x_{id} - x_{kd}) \cdot p$
      end for
      $h_{cd} = temp$
    end for
    $\boldsymbol{h}_c = \boldsymbol{h}_c\ /\ N_{\omega_c}$
    $temp = \max(\mathbf{last} - \boldsymbol{h}_c)$
    last $= \boldsymbol{h}_c$
  until $temp > dif$
end for

4. Experiments

A database of road sign images for classifier performance evaluation has been acquired. It contains 1100 images from 45 road sign classes. Only the sign boards, not whole traffic scene images have been collected. The image size varies from 15 to 150 pixels and images are stored in 24-bit color coding. All images have been acquired by the Olympus *Camedia* digital camera under general illumination conditions. Images were divided into nine groups according to their shape and color combination. The following list contains a brief description, typical road sign and color combination for each sign group:

- G1 triangular warnings (e.g. *Danger*), (red, white, black)
- G2 circ. *Closed to all vehicles* and *One-way*, (red, white)
- G3 circ. prohibitions, *Speed limits*, (red, white, black)
- G4 circ. *No Stopping*, (red, blue)
- G5 circ. obligatory, *driving directions*, (blue, white)
- G6 upside triangle, *Give way*, (red, white)
- G7 octagon, *Stop! Major road ahead*, (red, white)
- G8 diamond, *Right of way*, (yellow,black, white)
- G9 square, *Pedestrian crossing*, (blue, black, white)

Additional testing images were generated from the original ones by random scaling from 15 to 150 pixels, random rotation by $\pm 5°$ and by adding Gaussian noise. Thus, the experimental database contains 4945 noisy road sign images from 45 classes in nine groups.

The feature computation process starts with HSV color segmentation. From the segmented image several binary images are generated using colors specific for the particular road sign group. Features are then computed on the binary images. For each dataset, 24 features have been used. The only exception is the group G2 (separation of *Close-to-all-vehicles* and *One-way* from other prohibition signs) where just 12 features have been computed on the white color in the segmented image. All the data were preprocessed by standardization. The same testing method has been used for all experiments. Each dataset was split randomly into ten parts. Nine of them were used for training and the remaining one for classifier testing. Ten such experiments were performed to

Algorithm 3: Sample rejection

```
for all patterns x_i^{ω_c} do
  work = abs(x − x_i^{ω_c})
  work = work ./h_c
  temp = 0
  for d = 1, ..., D do
    temp += work_d
    if temp > s_r then
      goto 12 // reject current sample
    end if
  end for
  classcontrib += exp(−temp)
end for
```

complete the rotation through the whole dataset. Estimated means of measured error rates and corresponding standard deviations of the mean estimates are given in the Table 1 for following six classifiers:

- *Laplace* – product kernel classifier with Laplace kernel, vector of smoothing parameters
- *Gauss* – product kernel classifier with Gaussian kernel, vector of smoothing parameters
- *mixture* – linear mixture of Gaussian probability densities, diagonal covariance matrix
- *ldc* – linear classifier assuming normal densities and equal covariances
- *qdc* – quadratic classifier assuming normal densities
- *k-NN* – nearest neighbor classifier ($k = 1$)

Number of component counts have been tested for Gaussian mixture classifier and the best result was given. As the estimation of full covariances caused numerical problems diagonal covariance matrices have been used instead. For all the experiments the individual feature selection method with Fisher criterion (Fukunaga, 1990) has been used. Features were sorted according to criterion values. Subsets with n-best features ($n = 2, 4, \ldots, D$, where D is the dataset feature count) were stored. The numbers in Table 1 are the best results attained by each classifier and the corresponding feature size.

From results it follows, that basically two different problem types exist in the road sign database. The first is a set of easily separable datasets G3, G4, G6, G7 and G9. On the other hand, there are more difficult problems like G1, G2, G5 and G8. The performance of the product kernel classifier is generally high. In the case of easily separable classes it behaves comparably to other classifiers. For difficult problems like groups G1 and G2 it gives much better results as it fits the structure of the data better than the other approaches. The Laplace kernel classifier, presented in this paper, gives comparable results to the classifier with the Gaussian kernel. Nevertheless, the training of the Laplace kernel classifier is six to ten times faster than the training of the Gaussian one, depending on the dataset. It is mainly caused by the faster convergence rate of the classifier with the Laplace kernel. Contrary to the k-NN classifier kernel classifiers weight the local distances by smoothing which is estimated from the data. For some sign groups (like G1 and G2) it can be an advantage to use the kernel approach. However, k-NN classifier performs better for other datasets like G3 or G9. The results of the mixture classifier depend on the quality of the supplied model and the amount of data at hand. The number of components for each class is given in advance and the model is then initialized by k-means clustering algorithm. If a large number of components is used the training procedure (EM algorithm) often runs into numerical problems.

5. Conclusion

The goal of the paper has been to show the behavior of the Laplace kernel classifier on the real world problem like the road sign recognition and to compare its performance with other methods. It was tested on nine datasets with noisy road sign images. Simple features computed on binary results of color segmentation have been used. It has been shown experimentally that the Laplace kernel classifier offers high performance even for the difficult problems. The advantages of the presented approach are fast computation and the efficient way of learning (estimation of smoothing factors) by the EM algorithm-based maximization of the cross-validated log-likelihood function. The kernel classifier uses the data itself for the construction of the probability density estimate. This makes the

Table 1
Experimental results – mean error rates and standard deviations of mean estimates in percent. The number of features where the best results have been reached is given in parentheses

Group	Classes	Samples	Laplace (%)	Gauss (%)	mixture (%)	ldc (%)	qdc (%)	knnc (%)
G1	17	1369	17.5 ± 0.4 (24)	18.2 ± 0.6 (14)	27.6 ± 1.4 (18)	28.2 ± 0.5 (24)	23.8 ± 1.3 (14)	20.8 ± 0.8 (8)
G2	3	720	2.4 ± 0.6 (6)	2.6 ± 0.5 (12)	9.0 ± 0.8 (6)	13.9 ± 0.9 (6)	15.3 ± 0.6 (8)	8.9 ± 0.9 (4)
G3	5	516	1.2 ± 0.3 (22)	1.6 ± 0.5 (24)	1.6 ± 0.4 (20)	2.0 ± 0.5 (24)	1.6 ± 0.3 (22)	0.7 ± 0.6 (12)
G4	2	222	0.6 ± 0.4 (18)	1.2 ± 0.7 (18)	0.9 ± 0.6 (22)	0.3 ± 0.3 (14)	0.3 ± 0.3 (16)	0.9 ± 0.6 (12)
G5	9	627	5.4 ± 0.8 (18)	5.3 ± 0.7 (18)	7.8 ± 0.9 (12)	5.1 ± 0.7 (12)	5.7 ± 0.8 (22)	6.3 ± 3.0 (20)
G6	2	557	0.7 ± 0.2 (10)	0.5 ± 0.2 (8)	4.2 ± 0.8 (16)	1.9 ± 0.3 (14)	2.4 ± 0.4 (12)	1.1 ± 0.4 (8)
G7	2	420	0.8 ± 0.5 (14)	0.9 ± 0.3 (14)	1.2 ± 0.6 (6)	0.9 ± 0.5 (14)	0.9 ± 0.5 (16)	1.9 ± 0.5 (6)
G8	2	216	4.0 ± 1.1 (12)	4.4 ± 1.1 (10)	6.6 ± 1.0 (20)	4.4 ± 0.9 (24)	4.0 ± 1.4 (24)	5.2 ± 1.0 (10)
G9	3	298	1.5 ± 0.5 (14)	1.3 ± 0.6 (14)	3.7 ± 1.3 (12)	0.9 ± 0.4 (18)	1.3 ± 0.4 (10)	0.9 ± 0.4 (6)

approach applicable to problems with small and multimodal data sets e.g. in the area of the road sign recognition. The disadvantage of kernel classifiers is that the whole training dataset is used for each computation of the probability density. The presented sample rejection method can reduce the amount of computation by rejecting useless samples from the processing.

Acknowledgements

This work has been partially supported by the Grant No. A2075608 and No. A2075606 of the Academy of Sciences and by the Grant of the Ministry of Education No.VS 96063 of the Czech Republic and the Complex research project of the Academy of Sciences of the Czech Republic No. K1075601.

References

Aldon, M.J., Pujas, P., 1995. Robust colour image segmentation. Seventh International Conference on Advanced Robotics, San Filiu de Guixols, Spain, September 20–22.

Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via EM algorithm. J. Royal Stat. Soc. 39, 1–38.

Devroye, L., Györfi, L., Lugosi, G., 1996. A Probabilistic Theory of Pattern Recognition. Springer, New York, ISBN 0-387-94618-7.

Duin, R.P.W., 1976. On the Choice of Smoothing Parameters for Parzen Estimators of Probability Density Functions. IEEE Transactions on Computing 25, 1175–1179.

Escalera, A., Moreno, L.E., Salichs, M.A., Armingol, J.M., 1997. Road Traffic Sign Detection and Classification. IEEE Transactions on Industrial Electronics 44 (6), 848–859.

Franke, U., Gavrilla, D., Görzig, S., Lindner, F., Paetzold, F., Wöhler, C., 1998. Autonomous Driving Goes Downtown. IEEE Journal of Intelligent Systems 13 (6).

Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition. Academic Press, ISBN 0-12-269851-7.

Lalonde, M., Li, Y., 1995. Road Sign Recognition. Technical report, Centre de recherche informatique de Montréal, Survey of the State of the Art for Sub-Project 2.4, CRIM/IIT.

Líbal, V., Paclík, P., Kovář, B., Mošna, P., Vlček, M., Zahradník, P., 1998. Road Sign Recognition System using TMS320C80. Proceedings of second European DSP Education and Research Conference, ESIEE, Paris, France.

Líbal, V., Paclík, P., Kovář, B., Zahradník, P., Vlček, M., 1997. Road Signs Recognition System. TI DSP Challenge, Praha.

Líbal, V., Zikmund, T., Paclík, P., Králík, M., Kovář, B., Zahradník, P., Vlček, M., 1996, Traffic Sign Identification and Automatic Template Generation. Proceedings of Workshop'96, ČVUT Praha, Part I, 185–186.

McLachlan, G., 1992. Discriminant Analysis and Statistical Pattern Recognition. Wiley, New York.

Nagel, H.H., Computer Vision for Support of Road Vehicles Drivers. Institut für Algorithmen und Kognitive Systeme, Uni.Karlsruhe, http://www-kogs.iitb.fhg.de/cveducat/Drivers/.

Paclík, P., 1998. Automatical Classification of Road Signs. Master thesis, Faculty of Transportation Sciences, Czech Technical Univeristy, Prague.

Paclík, P., Novovičová, J., 2000. Road sign classification without color information, ASCI 2000, Proceedings of Sixth Annual Conference of the Advanced School for Computing and Imaging, Lommel, Belgium, June 14–16, ASCI, Delft.

Priese, L., Klieber, J., Lakmann, R., Rehrmann, V., Schian, R., 1994. New results on traffic sign recognition. Proceedings of the Intelligent Vehicles Symposium. IEEE, Paris, October 24–26.

Sain, S.R., 1994, Adaptive Kernel Density Estimation. PhD thesis, Rice University, Houston, TX.