Detection of Traffic Signs Using Posterior Classifier Combination

Lucas Paletta*

Institute of Digital Image Processing, Joanneum Research
Wastiangasse 6, A-8010 Graz, Austria

Abstract

Mobile mapping of environment information from a moving platform plays an important role in the automatic acquisition of GIS (Geographic Information Systems). The extraction of railway infrastructure from video frames captured on a driving train requires a robust visual object detection system that provides both high localization accuracy and the capability to cope with uncertain information. This work presents a radial basis functions (RBF) neural network that models appearance based object recognition of traffic lights and railway signs within a probabilistic framework. A comparison of different classifier combination strategies demonstrates that a classifier prioritization scheme based on an information theoretic selection criterion provides the best recognition performance.

1. Introduction

Object detection plays an increasingly important role for the mobile mapping of environment information [14]. E.g., automatic extraction of traffic infrastructure from videos captured on a mobile platform enables to establish a most realistic model of the environment which may serve multiple purposes such as inspection of thousands of miles of traffic systems, or the generation of simulators for the training of train drivers [8]. Furthermore, with the integration of visual information and georeference signals based on GPS/INS, the acquisition of complex environment information into GIS (Geographic Information Systems) becomes both more accurate and feasible [3].

The original contribution of this work is to introduce a robust method for railway sign detection as essential part of an integrated system for automatic detection of railway infrastructure from image sequences [8]. The video frames are first segmented into regions of interest by exploiting a-priori knowledge about the visual scene, and classified using a posterior radial basis functions (RBF) network [12] that has been trained from real imagery. Appropriate classifier combination provides then robust detection across different sign classes.

In a similar framework, several methods have been proposed for the recognition of road signs. They rely on color classification [11], which is dependent on daylight and weather conditions, depend on the recognition of shape classes [6], or optimize a neural classifier in an evolutionary adaptation process [4]. The presented work demonstrates a both robust and rapid method for the application of railway sign detection. The symbolic description of the railway infrastructure is then georeferenced using GPS and INS signals.

The paper gives an outline of the probabilistic object detection methodology and describes preliminary results.

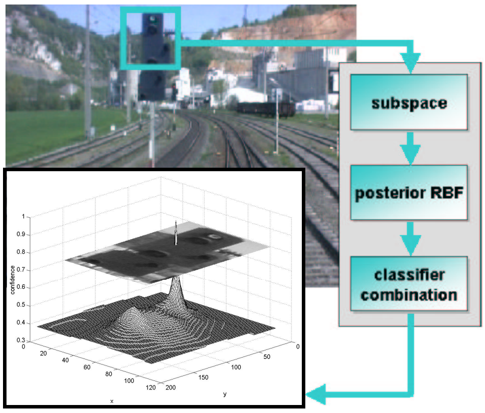


Figure 1. Object detection of railway signs. Local image patterns are projected to subspace, mapped by posterior neural networks (RBF) for a probabilistic interpretation, and combined for a final MAP decision.

2. Probabilistic object detection

The presented probabilistic railway sign detection first applies a dimensionality reduction on the sensory input space and then uses a neural network classifier that provides posterior distributions in the subspace to enable probabilis-

*lucas.paletta@joanneum.at

1051-4651/02 $17.00 (c) 2002 IEEE

tic reasoning on the object hypotheses. Appropriate classifier combination achieves then the most promising recognition results.

Appearance based object recognition The detection process is based on a recognition module operating on local image patterns which are successively extracted from the image (Fig. 1), e.g. in a row-wise manner. Appearance based object models consist of a collection of raw sensor footprints combining effects of shape and reflectance [7]. Instead of storing high-dimensional pixel patterns $\mathbf{x}_i$, the sensor vector can be transformed, e.g. by principal component analysis (PCA), to a low-dimensional representation $\mathbf{g}_i$ in a subspace called *eigenspace*. It captures the maximum variations in the presented data set whereas distances are a measure of image correlation. Recognition is supported by the property that close points in subspace correspond to similar object appearances [2].

Posterior RBF networks Object representations that implicitly model the uncertainty in eigenspace must consider estimates of the data density. Moghaddam & Pentland [7] applied a Bayesian framework on PCA descriptions with density estimations provided by unsupervised clustering. Approaches assuming *closed world* perception refine the confidence estimates by supervised learning. E.g., Ranganath & Arun [12] described how to combine eigenspace features and a RBF network for face recognition. The present system uses this concept under definition of a *rejection class* w.r.t. background to enable a closed world interpretation, Table 1).

RBF networks [12] consist of an input layer, a hidden layer of basis functions, and an output layer of linear activation units. The RBF network mapping for *k*-th output node y_k, $k = 1 \ldots \Omega$, Ω is the number of objects, operating on the input feature vector $\mathbf{g}_i$, is described by

$$y_k(\mathbf{g}_i) = \sum_{j=1}^{M} w_{kj} \phi_j(\mathbf{g}_i), \qquad (1)$$

with M basis units ϕ_j connected to the outputs by weights $[\mathbf{W}]_{kj}$. The Gaussian is a common choice of basis function, parametrized by its center vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$. Training of the free parameters is preferably separated into two phases. First, the basis centers $\boldsymbol{\mu}_j$ are identified with the centers of representative data clusters, and the output weights $[\mathbf{W}]_{kj}$ are determined thereafter under minimization of a sum-of-squares error function.

For a probabilistic interpretation of the input data, the outputs of the RBF network can be tuned to have a precise interpretation in terms of the posterior probabilities [13]. To evaluate any test data, the feature vector $\mathbf{g}_i$ is fed to the RBF network and mapped to the output values y_k for a posterior estimate,

$$\hat{P}(o_k|\mathbf{g}_i) = \alpha y_k(\mathbf{g}_i), \qquad (2)$$

α is a normalizing constant. $\mathbf{g}_i$ is then labelled by the *maximum a posteriori* (MAP) object hypothesis $o_{MAP} = \arg\max_{k=1}^{\Omega} \hat{P}(o_k|\mathbf{g}_i)$.

Classifier combination The posterior beliefs in object hypotheses o_k can be obtained from different interpretations $\mathbf{g}_{i,n}$, $n = 1..N$, (e.g., different spectral channels $\mathbf{x}_{i,R}, \mathbf{x}_{i,G}, \mathbf{x}_{i,B}, n \in \{R, G, B\}$) of multispectral sample $\mathbf{x}_i$ and are consequently updated with additional evidence. Combination of classifiers [5] are used for decision making by combining the individual opinions to derive consensus decisions with potentially more robust performance results.

We particularly investigate the *product rule* on individual posteriors, which represents a severe rule of classifier output fusion since a single recognition engine can inhibit the global interpretation [5]. $P(o_k|\mathbf{g}_{i,1}, \ldots, \mathbf{g}_{i,N})$ represents the joint probability distribution of the measurements extracted by the classifiers. Assuming conditionally statistical independence of the used representations $\mathbf{g}_{i,n}$, one obtains [10, 5]

$$P(o_k|\mathbf{g}_{i,1}, \ldots, \mathbf{g}_{i,N}) = \alpha P(o_k) \prod_{n=1}^{N} P(\mathbf{g}_{i,n}|o_k), \qquad (3)$$

with priors $P(o_k)$ and normalizing factor α. The product rule produces a fused decision making with the price of being sensitive to statistical outliers, $o_{prod} = \arg\max_{k=1}^{\Omega} P(o_k|\mathbf{g}_{i,1}, \ldots, \mathbf{g}_{i,N})$ [5].

Alternatively, the *max rule* determines a classification by the global MAP object hypothesis by,

$$o_{max} = \arg\max_{k=1}^{\Omega} \max_{n=1}^{N} P(o_k|\mathbf{g}_{i,n}). \qquad (4)$$

The Shannon entropy of the probability distribution from monospectral data,

$$H(o|\mathbf{g}_{i,n}) = -\sum_{k=1}^{\Omega} P(o_k|\mathbf{g}_{i,n}) \log P(o_k|\mathbf{g}_{i,n}), \qquad (5)$$

provides an *information theoretic criterion* to decide for the MAP hypothesis o_{MAP} of a single classifier from source n_{ent} by

$$n_{ent} = \arg\min_{n=1}^{N} H(o|\mathbf{g}_{i,n}), \qquad (6)$$

and therefore selects the MAP hypothesis with the most incisive confidence peak in the distribution.

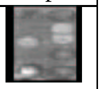
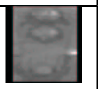
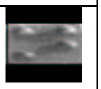
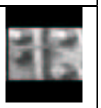
class	symbol	sample
Hauptsignal-HS		
Hauptsignal (back)-HSb		
Vorsignal-VS		
Vorsignal (back) -VSb		
Fahrleitungssignal-FS		
Geschwindigkeitstafel-GT		
Signalnachahmer-SNA		

Table 1. Object classes used in the experiments.

Detection process Detection is performed by a scan over the image in search for locations which produce a high confidence response of the posterior classifier (Fig. 1). The image content is masked by predefined masks that support the evaluation of object relevant information (Table 1). The masked image is scaled to predefined size and fed to the RBF classifier. Each specific mask provides a corresponding probability distribution over object hypotheses and a corresponding MAP decision. An appropriately masked image is then expected to provide a corresponding largest confidence value.

3. Experiments

For the recognition experiments, object templates were manually extracted from the video frames captured on a regular train of the Austrian Federal Railways (ÖBB). A total of 728 samples were generated from different kinds of environment and illumination conditions to provide a representative profile for a feasibility study on a restricted set of 7 most relevant sign objects (Table 1).

Each local pattern sample $\mathbf{x}_i$ was assigned to different object masks according to its associated object classes (Table 1). All but the image masked by its associated object outline were then labelled *background* patterns. Images $\mathbf{x}_{i,n}$ were normalized, scaled to size 100 × 100, and projected to the eigenspace of dimension 20 to $\mathbf{g}_{i,n}(\mathbf{x}_{i,n})$ (Section 2). The basis units of the neural classifier were first determined using the EM method [1], and the output weights were adjusted according to Section 2 to assure a probabilistic representation of output values.

	train		test	
feature space	$\mu[\%]$	$\sigma[\%]$	$\mu[\%]$	$\sigma[\%]$
R	93.7	1.1	87.8	2.3
G	93.3	0.9	86.4	2.0
B	94.9	1.2	81.9	2.4
RGB	95.3	2.3	84.2	3.2
RGB-prod	92.1	1.4	85.1	2.2
RGB-max	96.1	1.3	88.7	1.9
RGB-ent	96.1	1.1	89.1	2.0

Table 2. Recognition rates using posterior RBF networks w.r.t. different training data and decision fusion methods (section 2); μ=mean, σ=stdev.

class			test	
Nr.	label	*samples*	$\mu[\%]$	$\sigma[\%]$
1	HS	94	85.0	2.7
2	HSb	44	75.0	3.4
3	VS	119	100.0	0.0
4	VSb	61	87.6	2.9
5	FLS	10	100.0	0.0
6	GT	48	83.3	3.1
7	SNA	22	75.0	3.5
8	BGD	330	91.3	2.6

Table 3. Recognition rates using posterior RBF networks w.r.t. specific object classes (see Table 1, BGD (*background*)); μ=mean, σ=stdev.

For an evaluation of the complete sample set, the classification of individual R, G, and B channels was compared with various classifier combination stategies (Table 2) applying 5-fold cross-validation on the data. Note that due to distortions by DV compression, the transformation to HSI space was ignored. While monospectral classification acquires robust recognition rates (e.g., 87.8 %, R channel), performance decreases using the integrated (*RGB*) feature space. Fusion under the *product rule* [5] (Eq. 3, *RGB-prod*) on confidences of single channels R, G, B is experienced to provide inferior results. Increased performance is gained applying the *max rule* (Eq. 4, *RGB-max*). Eventually, the best performance (≈ 89.1%) was achieved using the *information theoretic selection* criterion on the single channel classifiers (Eq. 6, *RGB-ent*) used as basis for a MAP decision. Table 3 outlines the recognition rates with reference to individual object classes (*RGB-ent*). As a basis for a successful detection of traffic signs, the background informa-

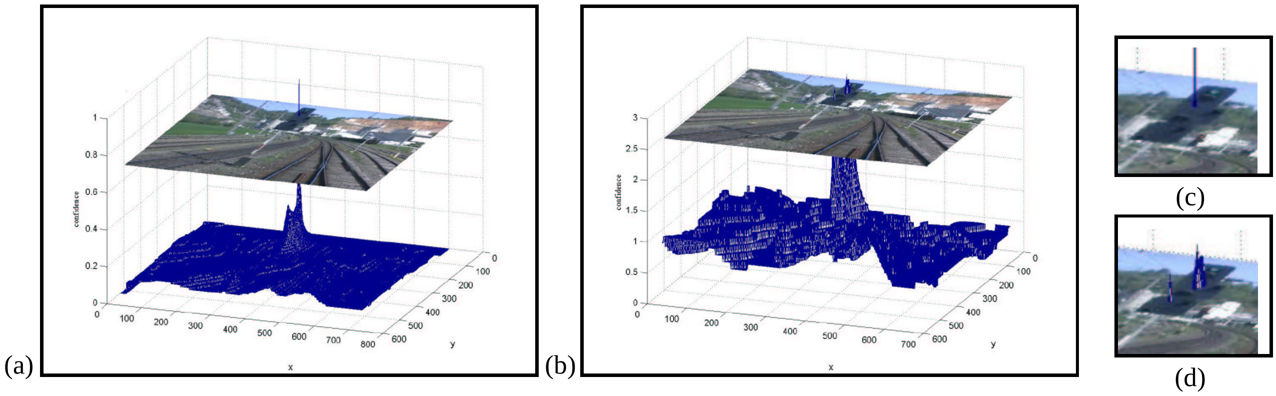


Figure 2. Object detection by probabilistic matching, depicting confidence responses w.r.t. (a,c) unambiguous and (b,d) ambiguous object hypotheses.

tion appears to be well discriminated ($> 91\%$ recognition rate) from sign information.

The detection process is performed by a pixel-wise scan of the neural classifier over the ROIs (complete image in Fig. 2), a process which results in a confidence landscape for each object hypothesis, with well-defined peaks experienced at the object's center position. While MAP confidences may provide a unique confidence maximum for precise localizations (Fig. 2a,c), the analysis with respect to a particular hypothesis may also result in ambiguous mappings (Fig. 2b,d). The presented railway sign detection system will be extended by probabilistic reasoning on the frame specific interpetations in order to resolve the ambiguities over an image sequence [9].

4 Conclusions

For vision based mobile mapping of railway infrastructure, it is essential to provide a robust and accurate object detection scheme that bases the decision making on uncertainty calculi such as probabilistic reasoning. The presented classifier combination strategy takes advantage of an information theoretic criterion on the selection of the decision maker to provide the best recognition performance within the experiments.

Future work is considered to focus on exploiting dynamic information from tracked visual patterns, such as to take into account the temporal context [9] within a sequence of video frames.

References

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B*, 39(1):1–38, 1977.

[2] S. Edelman. *Representation and Recognition in Vision*. MIT Press, Cambridge, MA, 1999.

[3] N. El-Sheimy and M. Lavigne. 3-D GIS data acquisition using GPS/INS/video mobile mapping system. In *Proc. International Symposium on Kinematic Systems in Geodesy, Geomatics and Navigation*. Vienna, Austria, 1998.

[4] S.-H. Hsu and C.-L. Huang. Road sign detection and recognition using matching pursuit method. *Image and Vision Computing*, 19:119–129, 2001.

[5] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on PAMI*, 20(3):226–239, 1998.

[6] J. Miura, T. Kanda, and Y. Shirai. An active vision system for real-time traffic sign recognition. In *Proc. IEEE International Conference on Intelligent Transportation Systems*. Dearborn, MI, USA, 2000.

[7] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on PAMI*, 19(7):696–710, 1997.

[8] L. Paletta, G. Paar, and A. Wimmer. Mobile visual detection of traffic infrastructure. In *Proc. IEEE International Conference on Intelligent Transportation Systems*, pages 616–621, Oakland, CA, 2001.

[9] L. Paletta, M. Prantl, and A. Pinz. Learning temporal context in active object recognition using Bayesian analysis. In *Proc. International Conference on Pattern Recognition*, pages 695–699, 2000.

[10] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Francisco, CA, 1988.

[11] L. Priese, R. Lakmann, and V. Rehrmann. Ideogram identification in a realtime traffic sign recognition system. In *Proc. IEEE International Conference on Intelligent Vehicles*, pages 310–314. Detroit, MI, 1995.

[12] S. Ranganath and K. Arun. Face recognition using transform features and neural networks. *Pattern Recognition*, 30(10):1615–1622, 1997.

[13] M. D. Richard and R. P. Lippmann. Neural network classifiers estimate Bayesian *a posteriori* probabilities. *Neural Computation*, 3(4):461–483, 1991.

[14] C. Tao, M. A. Chapman, and N. El-Sheimy. Towards automated processing of mobile mapping image sequences. In *Proc. 2nd International Workshop on Mobile Mapping Technology*, pages 2–5–1 – 2–5–10. Bangkok, Thailand, 1999.