A review of different object recognition methods for the application in driver assistance systems

Andreas Laika[1,2)] and Walter Stechele[2)]
[1)] BMW Group Forschung und Technik
Hanauer Straße 46, D-80992 München
[2)] Lehrstuhl Integrierte Systeme, Technische Universität München
Arcisstrasse 21, D-80333 München
Email: andreas.laika@bmw.de, walter.stechele@tum.de

***Abstract*— Algorithms in the field of driver assistance have been limited by their requirement for real time in the initial phase of their development. However, as computing power is increasing steadily, new possibilities arise. With focus on this situation a review is presented not defined by its designated field of application in driver assistance systems, but rather by the methods in use, namely video-based object recognition using machine learning. Recent methods are compared in a highly summarized table using criteria such as recognition rate, computational requirements or number of training samples required. Concluding their potential use in driver assistance is discussed.**

I. Introduction

A. Overview

Video-based object recognition has many applications in different fields such as image retrieval, surveillance systems and driver assistance systems to name just some. In all these fields the approach is similar: After some initial preprocessing features are extracted, which are then used for recognition. The different applications have different requirements on e.g. the data used (single images versus video), the additional information available, the expectations on reliability, and the requirements on computing time. Recognition systems are built according to these requirements. In driver assistance systems e.g. the timing constraints are strict. A frame has to be processed within 40 ms in order not to be dropped. Computational power is however growing steadily with every new generation of CPUs, so the requirement can be met for increasingly complex algorithms. Additionally suitable hardware acceleration is also becoming more available (see e.g. [12] and [13]). In a way constraints on computational complexity have become less strict, since more instructions can now be executed in a given time, than in the past.

B. Motivation

Bearing that fact in mind it seems to be worthwhile to compare methods of different fields. In order not to go beyond the scope of this paper we limit ourselves to object recognition tasks in the field of driver assistance. The motivation for this comparison is twofold: For one we want to list different methods in a table for easier comparison, for the other it is our objective to point out methods new to the application in driver assistance.

II. Scope of the Paper

In this paper we first discuss driver assistance systems. After an overview on some important aspects, different applications are presented to provide an impression on the state of the art in driver assistance systems. Similarly, we will talk about other fields of application, first providing some introductory words, then presenting examples of applications. In Table I all these methods are listed for better comparison. In one part we describe the processing steps, namely what algorithms are used for feature extraction and for classification. The other part is on how the overall system performs. One important criterion is of course the recognition rate. For the application in driver assistance systems timing is important and so computing cost is another item to be listed. Finally, the number of classes and the amount of samples used for training is another item on the list. Presenting only papers, which mention all these items, would confine us too much, however. So sometimes not all these items can be listed.

III. State of the Art

A. Driver assistance systems

Machine vision has been used in driver assistance and the related field of autonomous vehicles for quite some time. A review of this earlier development can be found in [14]. Reviews of more recent developments are presented in [15] and [16]. In contrast to many other applications here optical cameras are not the only sensors to be used. Usually additional information from e.g. Radar or Lidar is available. Stereo-based methods and methods using optical flow are also often available. This helps in the task of detecting different objects on and near the road like pedestrians, cars, trucks and road signs. On the other hand this means that there is less appeal for complex machine vision algorithms especially when they are also harder to implement due to computational constraints. There are a lot of methods relying on a-prioi knowledge in the form of a model, making use of the specific properties of that object, like a combination of edges or a template which the potential vehicle is matched against. Statistical pattern recognition methods are also used. Often they are employed in a two-step detection scheme, where the first step generates promising hypotheses (hypothesis generation: HG) in form of a region of

Name, Title and Reference of evaluated Method	Criteria of Evaluation				
	Number of classes and training samples	Recognition Rate	Computing Cost	Method used for Feature Extraction	Machine Learning Algorithm used
Kato and Ninomiya: „Preceding vehicle recognition based on learning from sample images“ in [1]	2 classes: 5000 vehicle and 5000 non-vehicle samples to train	Recognition rate: 96%-98%	Recognition: 0.5 - 1.7 ms @ PIII 800MHz	not further specified	MC-MQDF - Linear Classifier
Sun e.a.: „Monocular pre-crash vehicle detection: Features and classifiers“ in [2]	2 classes: 1051 vehicle, 1051 non-vehicle samples to train	Error rate: 3.8%-9.1%	Recognition: 100 ms @ PIII 1133MHz	PCA, wavelet, truncated / quantized wavelet, Gabor features, and combinations of those	Neural Networks & Support Vector Machines
Handmann, Kalinke, e.a.: „An image processing system for driver assistance“ in [3]	2 classes: vehicle / non vehicle; fully automatic processing	not specified	Recognition: $\approx$ 50 ms @ DEC Alpha 500MHz	features from shape, local symmetry, texture, shadows and color	LOC-classifier, Hausdorff distance classifier
Vailaya, Figueiredo e.a.: „Image classification for content-based indexing“ in [4]	2 classes: about 2500 samples to train	Accuracy: 88-95 %	not specified	10 x 10 sub-block color moments in LUV space, edge direction histograms	Bayesian Classifier with Vector Quantization for Density Estimation
Djordjevic, Izquierdo: „Empirical Analysis of Descriptor Spaces and Metrics for Image Classifcation“ in [5]	3 classes 12000 samples to train	not specified	not specified	MPEG-7 descriptors	Clustering with a variation of K-Median (PAM)
Viola and Jones: „Robust Real-time Object Detection“ in [6]	2 classes: face vs. non-face 4916 samples to train	Recognition rate: 78.3%-98%	Recognition: 0.07 s Training: $\approx$ 5 min @ PIII 700MHz	Haar-like features	AdaBoost
Huang and LeCun: „Large-scale Learning with SVM and Convolutional for Generic Object Categorization“ in [7]	6 classes: 24300 samples to train	Error rate: 5.9%-10.4%	Recognition: 0.03 - 0.95 s Training: 3 - 654 s @ 1GHz	not specified	Support Vector Machines, Convolutional Networks & combinations of both
Weber, Welling and Perona: „Unsupervised Learning of Models for Recognition“ in [8])	2 classes: 200 vehicle 200 non vehicle samples to train	not specified	Recognition: $<$ 1 s Training: 2 min @ PII 450MHz	Local features not further specified	Data clustering with the EM-Algorithm
Leibe e.a.: „Analyzing Appearance and Contour Based Methods for Object Categorization,“ in [9]	8 classes: 3280 images	Recognition rate: 64.85-86.4%	not specified	Color, texture, global shape (PCA), local shape	χ^2 measure for color / texture, shape matching
Torralba, Murphy and Freeman: „Sharing visual features for multiclass and multiview object detection“ in [10]	21 classes: about 50 samples/class	not specified	not specified	not specified	not specified
Opelt e.a.: „Generic object recognition with boosting“ in [11]	3 classes: 450 person, 350 bike 250 non bike / person samples to train	Recognition rate: 6.5-83.5%	not specified	intensity moments, SIFTs	AdaBoost

TABLE I

COMPARISON OF DIFFERENT OBJECT RECOGNITION SYSTEMS

interest and in a second step the pattern recognition algorithm verifies or rejects that hypothesis (hypothesis verification: HV). Kato and Ninomiya [1] recognize preceding vehicles taken from a conventional monochrome CCD camera without the help of additional systems like Radar or Lidar. They skip the HV by randomly selecting a region of interest (ROI), which is then down sampled to a normalized size. Feature extraction however is not specified further. These features are then used for training a linear classifier called MC-MQDF (multi clustered modified quadratic discriminant function). With 10000 images quite a high number of samples are used for training. Plots showing classification rates of the system for different parameters and different training sets (passenger vehicles, commercial vehicles and motorcycles) are presented. Classification rates range within 96%-98%. Another approach was published quite recently by Sun ea. in [2] and [17]. Here a special high dynamics camera is used to detect the rear of vehicles. Feature extraction is emphasized more in this paper. In a multiscale approach edges are extracted. These features are then used to locate a region of interest, where the actual processing is done. Different feature extraction methods like principal component analysis, wavelets and Gabor-features are applied. For classification neural networks (NN) and support vector machines (SVM) are used. The computation of these

features has its price in terms of computational cost. Detection results take on average 100 ms to compute. A complete system to detect, track, and classify vehicles is presented by Handmann and Kalinke in [3]. It is important to note that the terms detection and classification are often not used according to a strict definition. In [3] as in some other papers the term detection refers to the HG step where ROIs are identified, while the term classification is used for the HV step where a classification between two categories (vehicle versus non-vehicle) is performed. Quite a number of feature extraction methods like local orientation coding, polygon approximation of contours, the use of local symmetry, texture analysis based on local image entropy, shadow analysis and color analysis are used. Several classification algorithms are presented as well. The paper gives an impression on how a complete system in a car can work. For computing times single processing steps are specified. Unfortunately, the complete processing time can only be estimated from these. Probably due to the use as an online-system the overall recognition rate is also not specified.
In driver assistance, classification is mostly used for two class problems - a vehicle class is tested against a non-vehicle class. So different objects of interest on the road (vehicles, trucks, pedestrians, traffic signs) would have to be treated with different systems. This is not only expensive in terms of computing time, but also lacks flexibility. Additionally, also often only one perspective e.g. front or rear view of a car is used, which poses the question whether other views can be covered adequately by such systems. Clearly, a system being able to cover multiple classes and multiple views would be desirable.

B. Image retrieval

Image retrieval has a quite diametrical objective. Here the content of images is generally not limited to a specific context, and additional information is often also not available. This makes the task of natural image classification a hard one. A general review on the field can be found in [18]. It first deals with features obtained from color, texture and shape, and then discusses concepts of application. To complement this review [19] presents a thorough comparison of existing image retrieval systems. The different descriptors used for these systems are listed and compared.
An example of classification for image retrieval is presented in [4], where 2-class and 3-class classifiers are concatenated to a classifier with altogether 5 classes: That classifier is first discriminating Indoor/Outdoor, then City/Landscape and finally Sunset/Forest/Mountain. It is important to note, that in this case classification is performed on the whole image, not on objects in the image. Another paper dealing with classification is [5]. Here MPEG-7 descriptors were used to classify three different kinds of animals: Elephants, tigers, and horses. The focus is however not on the classification, but rather on examining how low-level features are related to human perception of images. Research is focusing on the semantic gap between low level features and the concepts they relate to. So what can be learned from image retrieval systems for driver assistance systems? Systems used in image retrieval will certainly not be adapted for the use in driver assistance applications in the near future. Too many issues concerning the semantic gap are unresolved. Driver assistance clearly has the advantage of a less generic scope. Some aspects dealt with in image retrieval may however also provide valuable insights for driver assistance systems. These could benefit from new methods for image segmentation or feature extraction (like e.g. [20]). It is however beyond the scope of this paper to go into detail here, and so we leave it at that remark.

C. General research in computer vision

There are also some research endeavours, which are not linked closely to a specific field of application, and are hence aggregated here as general research in computer vision. A review helpful in this context can be found in [21]. In Section III-A we already mentioned the desirability of multi-class and multi-view detection, but there are some other issues worth mentioning as well. For one there is the distinction between discriminative versus generative classification. Until now all the methods mentioned here have been of discriminative nature. Discriminative means that a decision rule is determined which separates two or more classes in a feature space. Generative approaches on the other hand try to determine a model for the various classes, e.g. by determining the probability distribution for each class. Usually discriminative approaches tend to be faster to learn, while generative approaches tend to perform better at recognizing non-class members. Important to note is the definition of the term model. Driver assistance systems also often deal with „model-based“ approaches. There a model is manually built into the system, describing for example the geometrical relations of edges in an image. In contrast to that, an example of a „model-free“ approach would be to let that happen automatically in the training. Geometric relations are however not limited to the domain of manually designing models. Usually pattern recognition algorithms in computer vision do not care about how features are arranged (geometrically) to each other. There are however also methods considering those relations, by building a geometric model. Again, those models should not be confused with models designed by hand. A final issue worth mentioning is feature selection. Selecting proper features is imperative for a recognition system, and attention has to be paid to this topic, when designing the system. A system with a certain ability to select features automatically is hence desirable, especially when dealing with several classes where the exact properties of features may not be known in advance.
Such a system is impressively demonstrated by Viola & Jones in [6], where a rather simple feature extraction using Haar-like features is combined with a cascade of AdaBoost-classifiers, proving good capabilities in the selection of features. An approach for multi-class and multi-view detection is suggested by [10]. In order to reduce training time and the required training samples, features are shared among a total of 21 classes. Here also a boosting technique is used for classification. Computing time is however not specified further, and the

recognition rate also seems to be below other more established methods. In [8] a generative approach is presented which builds a geometric model using an expectation-maximization (EM) algorithm. Parts of objects are extracted, clustered and then used in a binary classifier. Another method making use of boosting is presented in [11]. Here training is just weakly supervised, meaning that training images are just labeled as containing an object of a certain class or not, but no position of the image would be specified. Similarly to [5], Leibe studies in [9] how powerful different features like texture, color or shape are for categorizing objects. The system uses a decision tree, discriminating among 8 classes. Important to note, besides the fact that contours seem to be suited best is the error handling capability of the system. When an error is made, a similar class is more likely to be chosen. This leads to a graceful degeneration. Finally in [7] it is shown, that also with a more conventional approach using a combination of SVMs and convolutional nets a multi-category and multi-view system can be implemented.

In this paragraph we have learned, that the task of multi-class object recognition, often referred to as generic object recognition, has been tackled. There are various approaches covering interesting aspects of pattern recognition (e.g. discriminative versus generative classification, learning of geometric relations and feature selection). Some of these are however still experimental.

IV. Conclusion

Machine vision algorithms are well established today in the field of driver assistance. However, mainly conventional methods for detecting only one class of vehicles on the road are used. Limitations with these methods may occur when new tasks like the detection of multiples classes and/or multiple views are to be tackled. A first step in this direction would be the adaptation of a successful system like the one from [6] to a driver assistance system. Besides multiclass recognition, also the aspect of feature selection is covered here. Other interesting aspects like e.g. sharing of features (see [10]) could then be examined in a second step. Looking at more conventional approaches, shape was considered being a relevant feature in various fields. So a recognition system making use of that feature is another point worth looking at more closely. In this context shape-extraction / segmentation may be challenging, so the use of additional information, like stereo images, would have to be considered. Of course, to be used in driver assistance systems a method should be established to a certain level and not every new aspect of research in object recognition can be looked at. Whether a method is suited can ultimately not be decided by studying the literature, but has to be done by implementation and testing. This will be subject to further work.

Acknowledgment

The research leading to this work was supported by BMW Research and Technology and the COST292 Action on Semantic Multimodal Analysis of Digital Media.

References

[1] Takeo Kato, Yoshiki Ninomiya, and Ichiro Masaki, "Preceding vehicle recognition based on learning from sample images," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 3, no. 4, pp. 252–260, Dec 2002.

[2] Zehang Sun, George Bebis, and Ronald Miller, "Monocular pre-crash vehicle detection: Features and classifiers," *IEEE transactions on image processing*, vol. 15, no. 7, pp. 2019–2034, Jul 2006.

[3] Uwe Handmann, Thomas Kalinke, Christos Tzomakas, Martin Werner, and Werner v. Seelen, "An image processing system for driver assistance," *Image and Vision Computing*, vol. 18, no. 5, pp. 367–376, Apr 2000.

[4] A. Vailaya, M.A.T. Figueiredo, A.K. Jain, and Hong-Jiang Zhang, "Image classification for content-based indexing," *Image Processing, IEEE Transactions on*, vol. 10, no. 1, pp. 117–130, Jan 2001.

[5] D Djordjevic and E Izquierdo, "Empirical Analysis of Descriptor Spaces and Metrics for Image Classifcation," in *6th International Workshop on Image Analysis for Multimedia Interactive Services*, Montreux, Switzerland, Apr 2005.

[6] Paul Viola and Michael Jones, "Robust Real-time Object Detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2001.

[7] Fu Jie Huang and Yann LeCun, "Large-scale Learning with SVM and Convolutional for Generic Object Categorization," *Proc. Computer Vision and Pattern Recognition Conference (CVPR06)*, vol. 1, pp. 284–291, 2006.

[8] M. Weber, M. Welling, and P. Perona, "Unsupervised Learning of Models for Recognition," in *Proc. 6th European Conference Computer Vision (ECCV)*, Dublin, Ireland, Jun 2000, vol. 1, pp. 18–32, Springer.

[9] Bastian Leibe and Bernt Schiele, "Analyzing Appearance and Contour Based Methods for Object Categorization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, Madison, WI, Jun 2003.

[10] A. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing visual features for multiclass and multiview object detection," *IEEE Transactions On Pattern Analysis and Machine Intelligence*, In press.

[11] A. Opelt, A. Pinz, M. Fussenegger, and P. Auer, "Generic object recognition with boosting," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 3, pp. 416–431, Mar 2006.

[12] Walter Stechele, "Video Processing using Reconfigurable Hardware Acceleration for Driver Assistance," in *Workshop on Future Trends in Automotive Electronics and Tool Integration at DATE 2006*, Munich, Mar 2006.

[13] Alain Greiner, Frederic Petrot, and Mathieu Carrier, "Mapping an obstacles detection, stereo vision-based, software application on a multi-processor system-on-chip," in *Intelligent Vehicles Symposium 2006*, Tokyo, Japan, Jun 2006.

[14] E.D. Dickmanns, "The development of machine vision for road vehicles in the last decade," in *Intelligent Vehicle Symposium, 2002. IEEE*, Jun 2002, vol. 1, pp. 268–281.

[15] M. Bertozzi, A. Broggi, M. Cellario, A. Fascioli, P. Lombardi, and M. Porta, "Artificial vision in road vehicles," *Proceedings of the IEEE*, vol. 90, no. 7, pp. 1258–1271, Jul 2002.

[16] Zehang Sun, George Bebis, and Ronald Miller, "On-road vehicle detection : A review," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 5, pp. 694–711, May 2006.

[17] Zehang Sun, G. Bebis, and R. Miller, "On-road vehicle detection using evolutionary Gabor filter optimization," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 6, no. 2, pp. 125–137, Jun 2005.

[18] Yong Rui, Thomas S. Huang, and Shih fu Chang, "Image Retrieval: Current Techniques, Promising Directions And Open Issues," *Journal of Visual Communication and Image Representation*, vol. 10, no. 4, pp. 39–62, 2000.

[19] Remco C. Veltkamp and Mirela Tanase, "Content-Based Image Retrieval Systems: A Survey," Technical report uu-cs, Utrecht University, 2000.

[20] Serkan Kiranyaz, Miguel Ferreira, and Moncef Gabbouj, "A Novel feature extraction method based on segmentation over edge field for multimedia indexing and retrieval," in *Proceedings of WIAMIS Workshop*, Montreux, Switzerland, Apr 2005, pp. 13–15.

[21] Anil K. Jain, Robert P. W. Duin, and Jianchang Mao, "Statistical Pattern Recognition: A Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.