ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 301-306

Segmentation-Free Detection of Overtaking Vehicles with a Two-Stage Time Delay Neural Network Classifier

Christian Wöhler and Jürgen Schürmann

DaimlerChrysler Research and Technology
PO Box 2360, D – 89081 Ulm, Germany

Joachim K. Anlauf

Rheinische Friedrich-Wilhelms-Universität
Institut für Informatik II
D – 53117 Bonn, Germany

Abstract. We propose an algorithm based on a time delay neural network (TDNN) with spatio-temporal receptive fields for segementation-free detection of overtaking vehicles on motorways. Our algorithm transforms the detection problem into a classification problem of strongly downscaled image sequences which serve as an input to the TDNN without a preliminary segmentation step. The TDNN classifier is followed by an additional classification stage to evaluate the TDNN output over time, which achieves a significant enhancement of the detection performance especially under difficult visibility conditions.

1. Introduction

In this contribution we present a time delay neural network (TDNN) based algorithm for segmentation-free detection of moving objects on a non-stationary background. Our application domain is autonomous driving on motorways essentially performed by following the vehicle driving ahead. To keep the distance constant, stereo vision [3] is employed. In this scenario, our algorithm is needed to detect an overtaking vehicle such that it can be focused on as a new partner to be followed at higher speed without a need to establish 3D information by stereo vision about a second object in the field of view.
Many approaches to the detection of moving objects rely on an analysis of the optical flow vector field. An application to vehicle detection on motorways is mentioned in [5]. Another system for vehicle detection is described in [4] in which the spatial distribution of local gradient features like edges and corners serves as an input to a neural network. Once an obstacle has been detected, the classical approach is then to track characteristic object features along successive images (see e. g. [2]) by performing a rather exhaustive template matching

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 301-306

algorithm. In recent contributions (see e. g. [1]), these general concepts are extended to more difficult visibility conditions.
In our approach, we use a time delay neural network with spatio-temporal receptive fields for classifying sequences of input images that may contain an overtaking vehicle. As this special TDNN concept is already described in detail in [7, 8], we will only give a rough description in Section 2. In Section 3, two methods to enhance the detection performance by temporal evaluation of the TDNN output are presented. A summary of the paper is given in Section 4.

2. The TDNN classifier for image sequences

The activation pattern of the three-dimensional (xyt) input layer of the TDNN corresponds to the pixel values of the input image sequence. To take into account the locality of the object features, we make use of the concept of *spatio-temporal receptive fields*. This means that each neuron in the second neuron layer is not connected to the complete input layer but only to a limited region of it, called its receptive field. The receptive fields act as spatio-temporal filters which are adapted during the training process, producing several filtered versions of the input sequence in the second network layer. Each of these filtered versions of the input sequence is represented by the activation pattern of a three-dimensional group of neurons, which we call a *branch* of the network. The *shared weights* principle is applied, which means that each layer 2 neuron of a certain branch is connected to its receptive field by the same configuration of weights. The actual recognition process is then carried out in a third and a fourth network layer. The network output is given by the activations $\{\omega_1, \omega_2\}$ of two output neurons, one for the vehicle and one for the garbage class. The TDNN is trained to the output $\omega_1 = 1$, $\omega_2 = 0$ for an overtaking vehicle and to $\omega_1 = 0$, $\omega_2 = 1$ for a garbage pattern. After training, an input pattern is said to belong to the overtaking vehicle class if $q\omega_1 > \omega_2$, otherwise it is classified as a garbage pattern. Varying the parameter q yields a rate of classification (ROC) curve illustrating the trade-off between the classification errors of the two classes. The training algorithm is a rather simple backpropagation-like online gradient descent method which is described in detail in [7].
The image data is produced by a normal video camera fixed behind the windscreen and directed slightly to the left, connected to a Power PC 603 equipped with a framegrabber delivering greyscale half frames of resolution 720×288 pixels. In the application scenario of detecting overtaking vehicles, the input image sequences are obtained by cropping a region of interest (ROI) sized 350×175 pixels out of the left half of each half frame. This ROI is then downscaled to 32×32 pixels. As an overtaking process takes about one second and the hardware is able to grab, crop, and scale four ROIs per second, four subsequent ROIs are ordered into an image sequence, respectively. As this is done at each time step, two subsequent image sequences overlap by three images.
The training set consists of 484 examples displaying overtaking vehicles and 2402 garbage patterns. A training example is said to display an overtaking

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 301-306

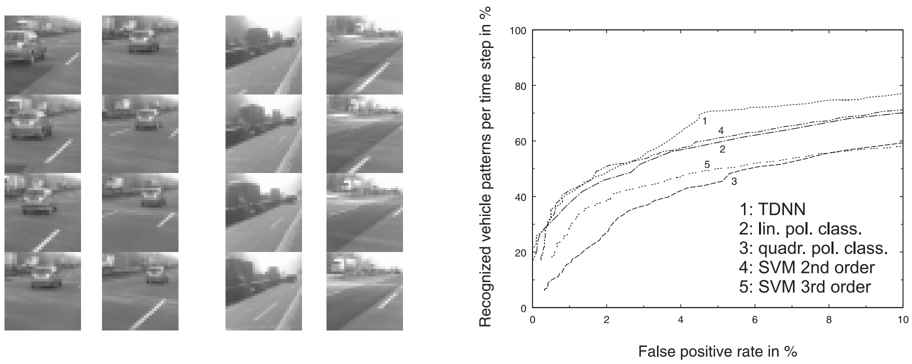


Figure 1: Left: Typical representatives of the overtaking vehicle class and of the garbage class. Right: Performance of the TDNN on the test set compared to more complex classifiers applied to the transformed feature vectors of reduced dimension in the second network layer (see also [9]).

vehicle if it has a velocity of at least 1 pixel per time step and a minimum size of 10 pixels on the first image of the sequence. The garbage examples contain many different views of empty motorway lanes including slopes and curves; moreover, they were collected under strongly varying weather conditions. Typical representatives of the two training classes are shown in Fig. 1. Our test set consists of 881 overtaking vehicle and 4369 garbage examples. It has to be rather large as for further examinations, it has to contain a statistically relevant total number of complete overtaking processes, which is 150 in our test set.
The TDNN configuration with spatio-temporal receptive fields of size $15\times15\times3$ pixels applied at an offset of $8\times8\times1$ pixels, 2 network branches and full connection between network layers 2 and 3 turned out to yield by far the best detection performance. The computation time of one classification process amounts to 10 ms on a Power PC 603. Results are shown in Fig. 1. The rather low recognition rate per time step has to be interpreted such that the beginning and the end of an overtaking process may quite often be missed, whereas a high rate of overtaking vehicles is detected at least once or twice during the overtaking process (Fig. 2). It turned out that applying more complicated classifier structures like higher-order polynomial classifiers and polynomial support vector machines (SVMs) to the transformed features in the second network layer as suggested in [9] yields no significantly improved results. Both for the polynomial classifiers and for the SVMs, an "overfitting" effect can be observed, i. e., the performance on the test set is decreasing with rising complexity of the classifier. This is probably due to systematic differences between the global properties of the training and the test set: the test set is a "difficult" one in the sense that it has been acquired independently of the training set several months later on different motorways and under different weather conditions. In the following section, however, we will discuss an alternative concept to enhance the detection performance.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 301-306

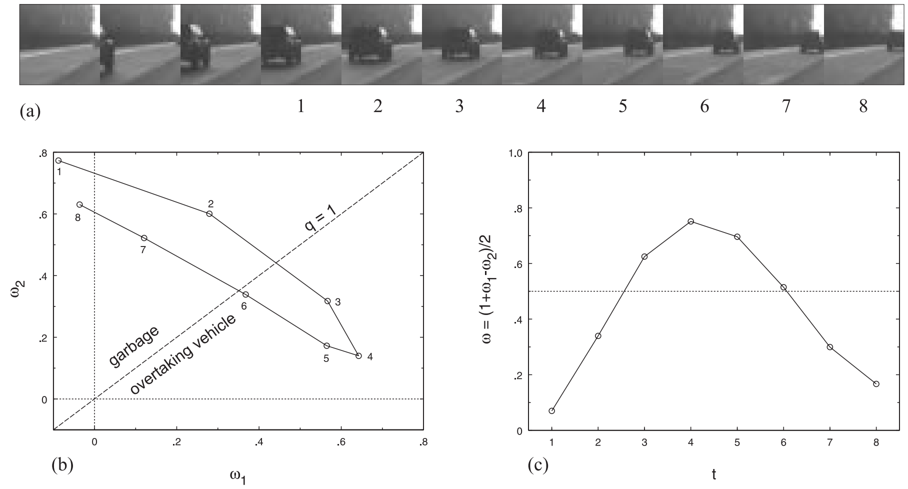
3. Evaluation of the TDNN output over time

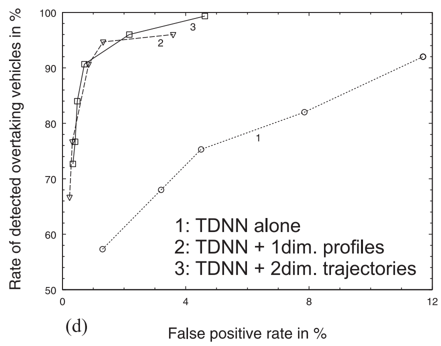
Evaluation of an averaged TDNN output: In the limit of an "ideal", i. e. infinite training set, the TDNN would act as an estimator for a posteriori probabilities and thus fulfill the condition $\omega_1 + \omega_2 = 1$ (cf. [6]). As intermediate patterns of vehicles just entering or exiting the scene are rather rare, systematic deviations from the line $\omega_1 + \omega_2 = 1$ are most distinct in the corresponding intermediate region of $\{\omega_1, \omega_2\}$ "output" space, leading to loops always performed in the same direction (Fig. 2b). We now neglect the two-dimensionality of the trajectory and only evaluate its projection $\omega = (1 + \omega_1 - \omega_2)/2$ on the line $\omega_1 + \omega_2 = 1$. We observed that the appearance of an impulse-like shape $\omega(t)$, $t = 1, \ldots, N_P$ more faithfully signals an overtaking vehicle than the $q\omega_1 > \omega_2$ condition, equivalent to $\omega > 1/(1 + q)$, that takes into account only one single network output; the occurrence of a loop or impulse may be evident even though the above conditions are not fulfilled.
The training set contains 886 impulse and 3073 non-impulse profiles, the test set which was derived from the same image sequences as the ones used in the previous section contains 1283 impulse and 3884 non-impulse profiles. The length of the profiles was chosen to be $N_P = 8$ (Fig. 2c). We classified the profiles with polynomial classifiers of degree 1 to 5 (cf. [6]). The error rate on the test set is significantly decreasing with an increasing degree of the classifier for degrees smaller than 5, such that the fourth-order classifier was used to compute the corresponding ROC curve in Fig. 2d.

Evaluation of the complete trajectory: In order to detect loop-like structures in output space as shown in Fig. 2b independent of their position and size which are largely variable, we regard the trajectory in output space as a sequence $c(t)$ of complex numbers at discrete time steps t with $c(t) = \omega_1(t) + i\omega_2(t)$. A Fourier transform (FFT) of an interval of N_P subsequent time steps should display strong high-frequency components in the case of a transition from the garbage towards the vehicle domain of the output space. To remove discontinuities at the borders of the interval, the sequence is mirrored according to $c(2N_P - t + 1) = c(t)$. The amplitude spectrum $A(k)$ of the sequence $c(t)$ nevertheless contains only N_P independent components as it can be shown that $A(N_P + 1) = 0$ and $A(k) = A(2N_P - k + 2)$ for $k = 2, \ldots, N_P$. It can be classified as representing loop or non-loop states. The phase spectrum turned out to contain no relevant information with respect to this classification task and was neglected. In this second approach, the performance of a linear polynomial classifier on the same training and test sequences as before could not be exceeded by higher-order polynomial classifiers or polynomial SVMs.

Result: Both proposed methods reduce the false positive rate by about an order of magnitude when the detection rate of complete overtaking processes is regarded (Fig. 2d). The test set contains 150 complete overtaking processes. Here, the false positive rate denotes the fraction between the time during which

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 301-306





Example: At time step 6, the sequence consisting of images 3, 4, 5, and 6 serves as an input to the TDNN.

Typical **garbage states**: 1, 8
Typical **overtaking vehicles**: 4, 5
Intermediate states: 2, 6

Figure 2: (a) Image sequence displaying an overtaking vehicle. (b) Trajectory of the TDNN output in output space. (c) Impulse profile $\omega(t)$ resulting from the same overtaking process. (d) Detection performance of the TDNN with and without the proposed second classification stages with respect to complete overtaking processes.

an overtaking vehicle is erroneously detected and the time during which in fact no overtaking vehicle is present. Although the evaluation of the complete two-dimensional trajectory gives no better results than an evaluation of the one-dimensional $\omega(t)$ profiles, significantly fewer classifier coefficients have to be adapted as a linear classifier is already sufficient. We could show that no loss in performance on the test set occurs for the second method when only every 10th training pattern is used; a procedure that leads to strong overfitting effects when applied to the fourth-order classifier necessary to successfully employ the first method. Furthermore, the direction in which the loop is performed might in principle be used to distinguish the direction of motion of the object in the scene.

4. Summary and conclusion

We propose a TDNN-based algorithm for the detection of moving objects on a differently moving background. It is applied to the detection of overtaking vehicles on motorways. Our algorithm transforms the detection problem into a classification problem of image sequences such that no preliminary segmentation stage is necessary. As the TDNN alone may fail to detect an overtaking vehicle if the contrast to the background is low or the visibility conditions are difficult, we suggest two ways of evaluating the TDNN output over time. The first one consists of classifying the temporal behaviour of an appropriate average of the output activations of the TDNN, the second one of classifying the amplitude spectrum of the complete two-dimensional trajectory in the output space. Both rather simple and time-efficient second classification stages reduce the false positive rate by about an order of magnitude especially under difficult visibility conditions.

Acknowledgement: The polynomial classifier and SVM results were obtained by applying the ClassLab toolbox written by Dr. Ulrich Kreßel of DaimlerChrysler Research and Technology.

References

[1] M. Betke, H. Nguyen. Highway Scene Analysis from a Moving Vehicle under Reduced Visibility Conditions. *IEEE Int. Conf. on Intelligent Vehicles*, pp. 131-136, Stuttgart, 1998.

[2] R. O. Duda, P. E. Hart. Pattern classification and scene analysis. Wiley-Interscience, 1973.

[3] U. Franke, I. Kutzbach. Fast Stereo based Object Detection for Stop&Go Traffic. *IEEE Int. Conf. on Intelligent Vehicles*, pp. 339-344, Tokyo, 1996.

[4] C. Goerick, D. Noll, M. Werner. Artificial Neural Networks in Real Time Car Detection and Tracking Applications. *Pattern Recognition Letters*, vol. 17, pp. 335-343, 1996.

[5] W. Krüger, S. Rössle, W. Enkelmann. Real-Time Estimation and Tracking of Optical Flow Vectors for Obstacle Detection. *IEEE Int. Conf. on Intelligent Vehicles*, pp. 304-309, Detroit, 1995.

[6] J. Schürmann. Pattern classification. Wiley-Interscience, 1996.

[7] C. Wöhler, J. K. Anlauf. A Time Delay Neural Network Algorithm for Estimating Image-pattern Shape and Motion. *Image and Vision Computing Journal*, in press.

[8] C. Wöhler, J. K. Anlauf, T. Pörtner, U. Franke. A Time Delay Neural Network Algorithm for Real-Time Pedestrian Recognition. *IEEE Int. Conf. on Intelligent Vehicles*, pp. 247-252, Stuttgart, 1998.

[9] C. Wöhler, U. Kreßel, J. Schürmann, J. K. Anlauf. Dimensionality Reduction by Local Processing. Submitted to *European Symposium on Artificial Neural Networks*, Bruges, 1999.