

Cellular neural networks for motion estimation and obstacle detection

D. Feiden and R. Tetzlaff

Institute of Applied Physics, Johann Wolfgang Goethe-University of Frankfurt a. M., Robert-Mayer-Str. 2-4, 60054 Frankfurt, Germany

Abstract. Obstacle detection is an important part of *Video Processing* because it is indispensable for a collision prevention of autonomously navigating moving objects. For example, vehicles driving without human guidance need a robust prediction of potential obstacles, like other vehicles or pedestrians. Most of the common approaches of obstacle detection so far use analytical and statistical methods like motion estimation or generation of maps.

In the first part of this contribution a *statistical algorithm for obstacle detection* in monocular video sequences is presented. The proposed procedure is based on a motion estimation and a planar world model which is appropriate to traffic scenes. The different processing steps of the statistical procedure are a feature extraction, a subsequent displacement vector estimation and a robust estimation of the motion parameters. Since the proposed procedure is composed of several processing steps, the error propagation of the successive steps often leads to inaccurate results.

In the second part of this contribution it is demonstrated, that the above mentioned problems can be efficiently overcome by using *Cellular Neural Networks* (CNN). It will be shown, that a *direct obstacle detection algorithm* can be easily performed, based only on CNN processing of the input images. Beside the enormous computing power of programmable CNN based devices, the proposed method is also very *robust* in comparison to the statistical method, because it shows much less sensibility to noisy inputs. Using the proposed approach of obstacle detection in planar worlds, a *real time processing* of large input images has been made possible.

1 Obstacle detection using statistical methods

In this first section a brief description of the problem and a statistical procedure (Feiden et al., 1999) of obstacle detec-

Correspondence to: D. Feiden
(Feiden@iap.uni-frankfurt.de)

tion in planar worlds using monocular image sequences will be given, whose performance will later be compared to the proposed CNN method. This summary will also demonstrate the complexity of a statistical obstacle detection approach.

1.1 Problem

Our studies concentrate on the use of *monocular image sequences*, which means that only a single camera is used for the recording of the video sequences.

Furthermore we are mainly interested in traffic scenes, where the condition that most world points are located on a plane is fulfilled. This concept is referred to as *planar world model*. Some images of such a traffic sequence are shown in Fig. 1, which form the raw data. Starting from these images, the goal of the statistical method (Feiden et al., 1999) is to establish a reliable motion estimation. Based on such a motion estimation, an obstacle detection can be realized afterwards (Carlsson and Eklundh, 1990).

1.2 The statistical procedure

In Fig. 2 the framework of the statistical obstacle recognition from monocular image sequences is illustrated in Feiden et al. (1999).

The first step of the procedure is to search for *prominent image areas*. Then the displacement of these areas in the subsequent image is calculated using the so called *displacement vector estimation*. After that, the determined list of displacement vectors can be used to calculate the three dimensional motion parameters and the normal vector of the plane¹. Finally, the calculated parameters are used to carry out a *motion compensation*, so that the result forms an image in which only three dimensional structures are visible and the texture of the plane representing the street has been removed. In the following the different steps of this procedure will be discussed in detail.

¹See “planar world model” in Sect. 1.1

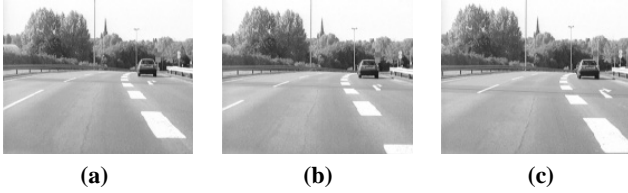


Fig. 1. Some pictures of a monocular image sequence.

1.3 Determination of prominent images

Since the calculation of correspondences between two images is a computationally very expensive operation, it is essential to restrict the matching process only to those regions where correspondences can be determined with sufficiently high confidence and accuracy. The scheme consists of two steps, a feature point extraction in *one* image and a subsequent block matching. A large number of methods have been proposed for extracting prominent image points; in Feiden et al. (1999) a slightly modified version of Förstner's and Gülch's feature point extraction method (Förstner and Gülch, 1987) is preferred, since this method is both theoretically conclusive and very powerful in practice. The basic idea of this operator is to compute estimates of gray scale gradients in a limited image area and to determine the eigenvalues of the covariance matrix of the computed set of gradients. The trace of the covariance matrix gives a measure of the texture content, whereas the isotropy of the gradient distribution is characterized by the relation of the eigenvalues of the covariance matrix. The result of this operation is the so called *prominence image*, in which corners are more prominent than edges, which is important for the subsequent displacement vector estimation.

1.4 Displacement vector estimation

A statistical displacement vector estimation is usually performed using some kind of *Block Matching* (Haralick and Shapiro, 1992). Block Matching is essentially a sort of correspondence analysis, where the gray scale differences between a reference block b_1 and a corresponding block b_2 are evaluated within a search area. Block b_2 is shifted through the whole search area and the sum of gray scale differences form a two dimensional error surface which should be zero at the right displacement position. Unfortunately, the error usually isn't zero anywhere due to noise effects. One simple possibility to determine the displacement vector is to take the position of the error minimum. Another improved method is described in Mester and Hötter (1995), where the gray scale difference residuals are converted into a local probability distribution and the expectation value and the covariance matrix of the displacement vector are evaluated subsequently. With this method an individual measure of confidence and accuracy for each single correspondence or displacement vector can be obtained.

1.5 Motion estimation and obstacle detection

The relation (Tekalp, 1995) between any corresponding points \mathbf{u} and \mathbf{v} of two successive images, with the components of the undisturbed displacement vector (v_{1p} and v_{2p}), the rotation matrix $\mathbf{R} = \{r_{ij}\}$ and the focal length f , is known to be

$$v_{1p} = \frac{u_1(r_{11} + t_1b_1) + u_2(r_{12} + t_1b_2) + f(r_{13} + t_1b_3)}{u_1(r_{31} + t_3b_1)/f + u_2(r_{32} + t_3b_2)/f + (r_{33} + t_3b_3)} \quad (1)$$

$$v_{2p} = \frac{u_1(r_{21} + t_2b_1) + u_2(r_{22} + t_2b_2) + f(r_{23} + t_2b_3)}{u_1(r_{31} + t_3b_1)/f + u_2(r_{32} + t_3b_2)/f + (r_{33} + t_3b_3)} \quad (2)$$

These equations are valid only if the measurement does not contain any errors. Unfortunately, the measured coordinates in the second image are superposed by additive errors, $v_1 = v_{1p} + z_1$ and $v_2 = v_{2p} + z_2$. The error vector \mathbf{z} is zero-mean and has a covariance matrix \mathbf{C} that is obtained from the measurement of the displacement vectors. Assuming that the measurement errors in the vectors \mathbf{v}_i follow a two-dimensional normal distribution with mean zero and covariance matrix \mathbf{C}_i , the resulting loss function $Q(\mathbf{p})$ has the form (Feiden et al., 1999) $Q(\mathbf{p}) = \sum_{i=1}^N |\mathbf{W}_i^T \mathbf{z}_i(\mathbf{p})|^2$, where \mathbf{W}_i is the root of the inverse covariance matrix \mathbf{C}_i^{-1} of \mathbf{z}_i , so that $\mathbf{C}_i^{-1} = \mathbf{W}_i \mathbf{W}_i^T$. Then the parameter vector \mathbf{p} can be determined by minimizing the loss function $Q(\mathbf{p})$.

For the final obstacle detection, motion compensated difference images can be used. The main idea of this procedure, which has been developed by Carlsson and Eklundh (Carlsson and Eklundh, 1990), is to transform every pixel of the actual image one time unit back, using the estimated parameters (see Fig. 3c). Then the difference image between the transformed image and the recorded image has to be calculated (see Fig. 3d).

If the parameters has been estimated correctly, and the image content is in accordance to the planar world model, the motion compensated difference image should appear in a homogeneous gray. However, real traffic sequences contain also three dimensional objects, which stand out in the motion compensated difference image. Finally, the obstacle detection can be carried out on the basis of this resulting image.

2 Direct obstacle detection using CNN

In their original paper (Chua and Yang, 1988), Chua and Yang proposed a single layer *Cellular Neural Network* (CNN) with linear coupled cells. According to our investigations, an obstacle detection in planar worlds is not feasible by using such a standard CNN. Therefore we focused on CNN with nonlinear weight functions.

2.1 Nonlinear CNN

In dropping the demand of linear coupled cells, nonlinear cell couplings can be characterized by polynomial functions.

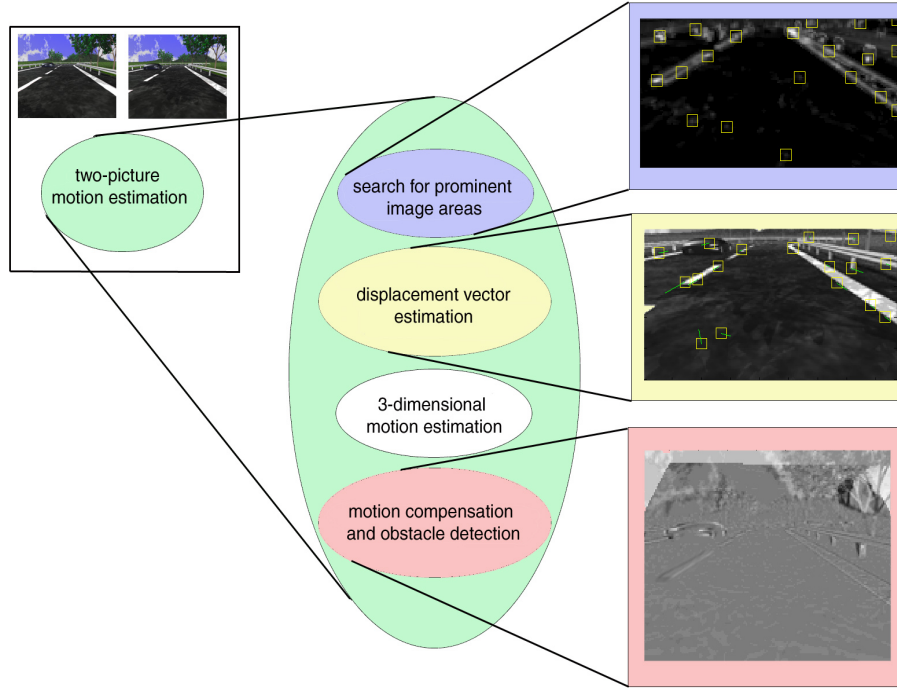


Fig. 2. The procedure of statistical obstacle detection in planar worlds.

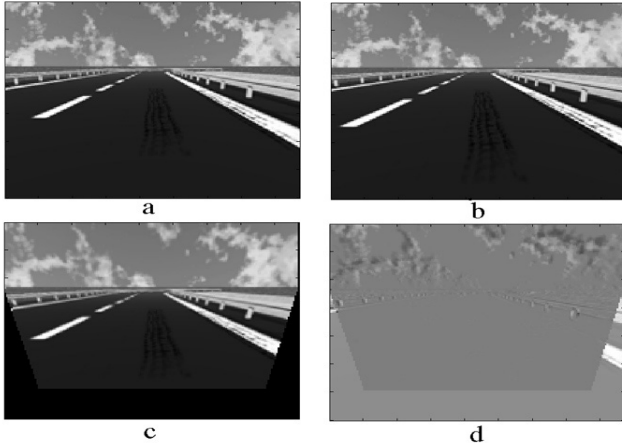


Fig. 3. Motion compensation and obstacle detection.

The dynamical behavior of a single layer, translation invariant CNN can be described by

$$C \frac{dx_i(t)}{dt} = -\frac{1}{R} x_i(t) + \sum_{j \in \mathcal{N}_i(k)} a_{i-j}(y_j(t)) + \sum_{j \in \mathcal{N}_i(k)} b_{i-j}(u_j(t)) + I \quad (3)$$

with $i = 1, \dots, N$,

where $a_{i-j}(y_j(t))$ and $b_{i-j}(u_j(t))$ are the polynomial feedback and feedforward functions, which are given by $a_{i-j}(y_j(t)) = \sum_{d=1}^D \alpha_{i-j,d} \cdot (y_j(t))^d$ and $b_{i-j}(u_j(t)) =$

$\sum_{d=1}^D \beta_{i-j,d} \cdot (u_j(t))^d$. The parameter D denotes the polynomial order, e.g. $D = 2$ describes a quadratic coupling function. The polynomial coefficients $\alpha_{i-j,d}$ and $\beta_{i-j,d}$ and the threshold value I form the parameters, which define the behavior of the CNN.

2.2 The CNN-procedure for direct obstacle detection

In Fig. 4 a few images of a synthetical image sequence are given, showing a ride over a textured plane on which three dimensional objects are located, which has been recorded by a moving camera. As in real traffic scenes, the motion direction and the viewing direction are identical.

The goal is to find a CNN that is capable to extract the three dimensional objects by presenting two images of such a sequence.

All of the image structures located on a plane should be erased by the CNN, so that only the 3D-objects remain as potential obstacles. It seems to be likely, that such a classification can only be performed on corner pixels (object corners as well as corners of texture), because only those pixels vary between two subsequent images. This is the reason why the procedure illustrated in Fig. 5 has been chosen for CNN based obstacle detection.

The first processing step is a simple edge extraction for both of the input images, hence a standard CNN can be used (Roska and Kék, 1999). In the following step the thresholding (Roska and Kék, 1999) of the resulting images have to be carried out. In Fig. 6 the different preprocessing steps are shown for a single input image.

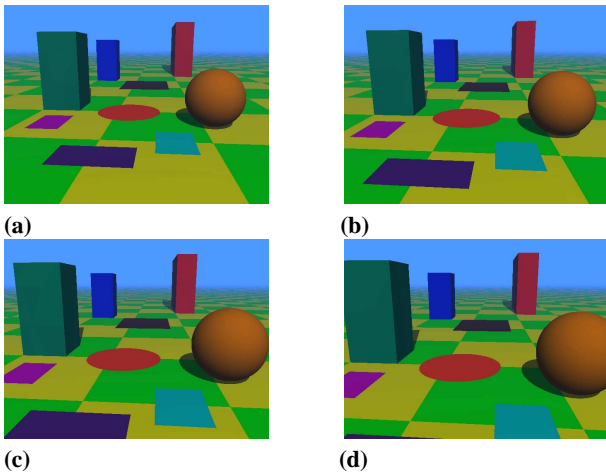


Fig. 4. A few pictures of a synthetically generated sequence.

some images of a video sequence showing a drive down the street

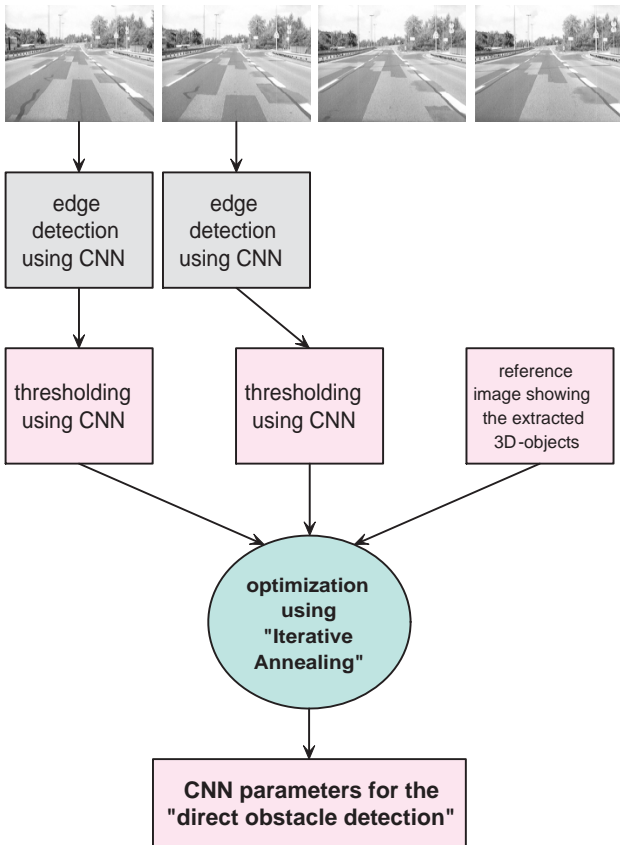


Fig. 5. The procedure of CNN based obstacle detection.

After these processing steps, the CNN parameters for the *direct obstacle detection* can be determined using the optimization procedure *Iterative Annealing* (Feiden and Tetzlaff, 2001). The two thresholded images together with the *reference image* containing only the edges of the three dimensional objects will be taken as input data for such an optimization.

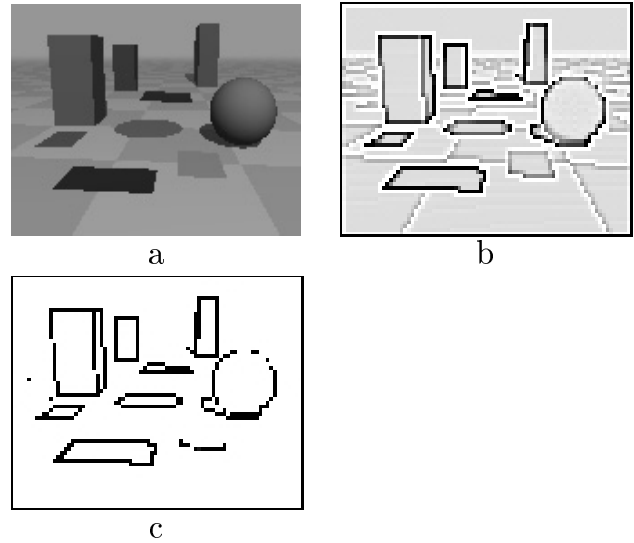


Fig. 6. Original image (a), CNN-generated edge image (b) and thresholding (c)

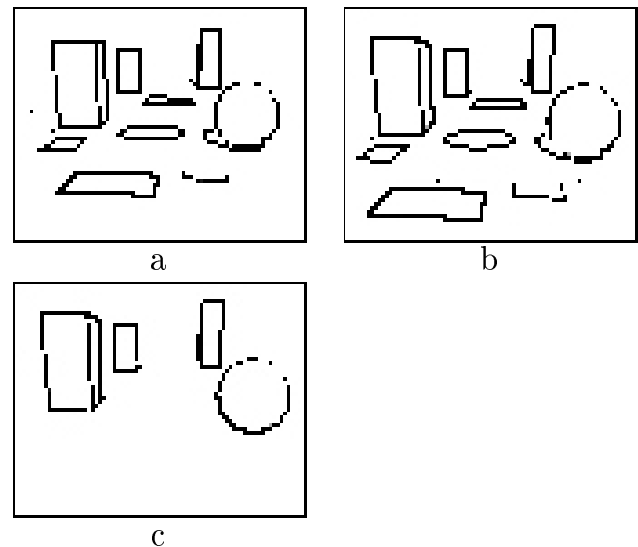


Fig. 7. The thresholded training images (a) and (b) and the reference image (c).

2.3 Results using synthetical image sequences

For the procedure shown in Fig. 5 a set of training patterns is obtained, that is presented in Fig. 7.

Using these training patterns (Fig. 7a–c) we determined a CNN which is capable to extract 3D-objects from the sequence. Our investigations yielded that a CNN with 5-by-5-neighborhood and polynomial cell coupling of degree 3 is appropriate for such a task. For the optimization the statistical procedure *Iterative Annealing* (Feiden and Tetzlaff, 2001) has been used. In Fig. 8b the CNN generated output is shown.

It can be observed that in the CNN output all textures on

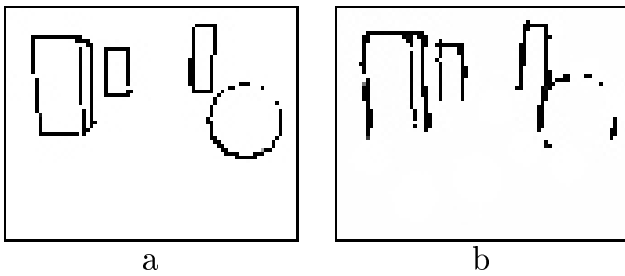


Fig. 8. Reference image (a) and CNN generated output image (b)

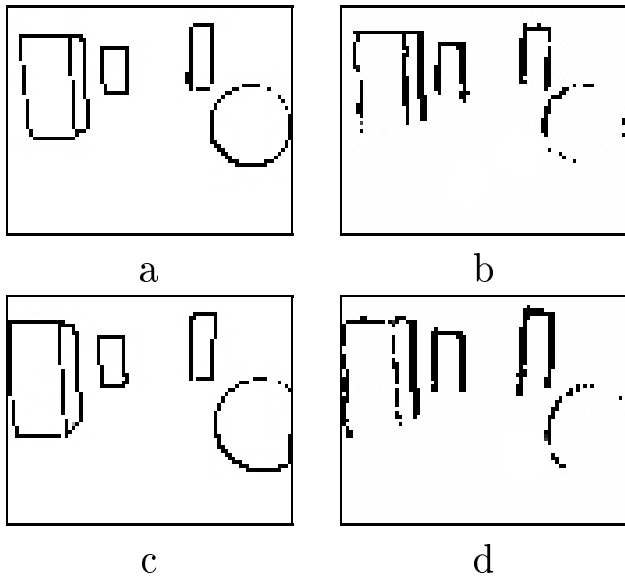


Fig. 9. The CNN generated output images (b) and (d) and the optimum results (a) and (c) for subsequent images of the synthetical sequence

the plane are removed and only the edges of the objects above that plane remain. It should also be noted that the lower edges of the objects also vanished. It is obvious that there is no possibility to discriminate edge pixels of texture on the plane from lower edge pixels of objects on that plane since both of them move in the same way.

In order to verify the determined CNN algorithms, other pictures of the video sequence have been processed as well by using the same CNN parameters. In Fig. 9 beside the “optimum” results² also the CNN output images are presented. Even in these untrained images the textures of the plane have been removed as well.

3 Conclusion

In Sect. 1 we have shown that statistical methods of obstacle detection are of high computational complexity. Furthermore, the robustness of the whole procedure depends sensitively on every individual processing step. E.g. a single outlier in the displacement vectors can disturb the motion estimation so strongly that the subsequent obstacle detection produces useless results. Furthermore we introduced in Sect. 2 a new method of obstacle detection using CNN. After two noncritical standard operations – edge extraction and thresholding – the obstacle detection can be performed directly by the CNN. The promising first results using synthetical as well as real image sequences demonstrate the high potential of CNN for the area of Computer Vision.

References

- Carlsson, S. and Eklundh, J.-O.: Object detection using model based prediction and motion parallax, Lect. Notes in Comp. Sc. 427, Springer, 1990.
- Chua, L. O. and Yang, L.: Cellular Neural Networks: Theory and Applications, IEEE Trans. on Neural Networks, 35,10, p. 1257–1272, 1988.
- Feiden, D., Mühlich, M., and Mester, R.: Robuste Bewegungsschätzung aus monokularen Bildsequenzen von planaren Welten, Förstner W.: Mustererkennung 99, 21. DAGM-Symposium, Bonn, 1999.
- Feiden, D. and Tetzlaff, R.: Iterative Annealing: A New Efficient Optimization Method for Cellular Neural Networks, Proc. of the Int. Conf. on Image Processing 2001, Thessaloniki, 2001.
- Förstner, W. and Gülch, E.: A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centers of Circular Features, Proc. of Int. Conf. on Fast Processing of Photogrammetric Data, Interlaken, 1987.
- Haralick, R. M. and Shapiro L. G.: Computer and Robot Vision, Addison Wesley, 1992.
- Mester, R. and Hötter, M.: Robust displacement vector estimation, Proc. 5th Int. Conf. on Image Processing and its Applications, UK pp. 168–172, 1995.
- Roska, T. and Kék, L.: CSL: CNN Software Library, Budapest, 1999.
- Tekalp, A. M.: Digital Video Processing, Prentice Hall, 1995.

²The binarized images of Fig. 7 with removed plane structures are referred to as “optimum” results.