Institut für Neuroinformatik
Ruhr-Universität Bochum

Internal Report 97–14

Image Processing of Dynamic Scenes

by

Werner von Seelen, Walter Gillner, Christian Goerick, Uwe Handmann, Thomas Kalinke, David Kastrup, Iris Leefken, Gesa Lorenz, Detlev Noll, Christos Tzomakas, Bernd Völpel, Martin Werner, Carsten Winkel

Ruhr-Universität Bochum
Institut für Neuroinformatik
44780 Bochum

IR-INI 97–14
July 1997
ISSN 0943-2752

Contents

Image Processing of Dynamic Scenes

Werner von Seelen, Walter Gillner, Christian Goerick, Uwe Handmann, Thomas Kalinke, David Kastrup, Iris Leefken, Gesa Lorenz, Detlev Noll, Christos Tzomakas, Bernd Völpel, Martin Werner, Carsten Winkel

Abstract

Systems for automated image analysis are useful for a variety of tasks and their importance is still increasing due to technological advances and an increase of social acceptance. The main focus of “Technical Image Processing of Dynamic Scenes” lies with the development of methods for the interpretation of images derived from various sensors. Apart from conventional visual images, this involves mainly X-ray and radar images. Taking into account the requirements of the various applications, suitable methods are derived.

Current projects are dealing with the analysis of traffic scenes (cf. section 1), detection of detonators when X-raying luggage (cf. section 2), and determination of type and expansion of oil pollution in maritime surveillance (cf. section 3).

1 Computer vision for driver assistance systems

Fully or partly autonomously guided vehicles, particularly for road-based traffic, pose high demands on the development of suitable algorithms due to the conditions imposed by natural environments. At the Institut für Neuroinformatik, methods for controlling vehicles by computer vision are developed in cooperation with several partners from the automobile industry.

Principal problems are caused by having a moving observer in predominantly natural surroundings classifying moving objects with a task determined operating frequency and estimating their positions. In particular, it is attempted to isolate traffic participants from video images and to attribute the obtained object hypotheses (e.g., object class, distance, velocity, danger potential with respect to the planned self-movement trajectory, etc).

The main focus of the project partners' work, in contrast, lies with lateral guidance of the vehicle and the corresponding estimation of road boundaries.

The great variety of different scenarios as well as the high degree of reliability necessary for the given task require an encompassing and flexible system architecture. The requirements concerning the reliability of the reached solution, the variety of geometric appearances of involved objects and that of environmental constraints of both deterministic as well as statistical nature necessitate a multitude of partial solutions based on different representations of the environment. Consequently, complexity and structure of the overall system have to be adaptable to increasing system complexity in order to allow for accommodation of additional modules without degeneration of already accomplished partial solutions. For this reason, even 'simple' applications are encumbered by considerations concerning the overall system architecture.

Basically, the overall system architecture can be divided into basic algorithms and fusion or integration algorithms. Basic methods are those delivering specific partial solutions under given constraints (e.g., environmental conditions). Results and application of the individual algorithms are not independent, resulting in an increase in redundancy making the overall system secure and reliable given a suitable coupling architecture. The necessary methods for fusion and integration ensure a flexible cooperation of the basic building blocks as well as the integrative derivation of results from them. In a similar vein, a sequentiell data transmission and system dynamics are necessary in order to build up an overall system with feedback stability solving complex tasks.

In the area of preprocessing, a multitude of different methods for initial segmentation, object tracking, and object classification has been developed in the context of current research. A few inherent tendencies appear remarkable.

- Previous work often was based on the use of higher features, meaning the generation of a sequence of features beginning at the iconic (image-based) side and continuing to the symbolic side. There are two main reasons:

  1. The historic rooting of image processing in material and surface inspection for quality control has lead to the existence of theoretically well-founded and practically tested algorithms.
  2. Symbolic features are commonly used for compact coding purposes, so that processed data amounts can be largely reduced for accommodating limited processing resources.

  The breath-taking evolution of processors has particularly alleviated the impact of this last constraint. In addition, it appears that particularly in the context of limited sensor resolution (i.e., in large distance regions) algorithms can be employed that rely on statistical measures of extensive 'early' (in the chain of processing) feature sets. These algorithms supplement the spectrum of methods explicit in more traditionally oriented algorithms.

- Often a formulation as an optimization problem can lead to implicitly robust solutions avoiding disadvantages of explicit methods (e.g., the correlation of model with image features, the correspondence problem). In this area as well the increase in available computational power has been aiding progress.

- Particularly in natural environments, flexible algorithms possessing a certain learning capability for input data driven adaption are preferably used.

At the Institut für Neuroinformatik, a number of algorithms are available, on the one hand for providing partial solutions, on the other hand for aiming at accomplishing the architecture of an overall system.

- Fourier Mellin and cepstrum analysis,
- binocular stereo-based methods,
- inverse perspective mapping,
- local orientation coding,
- use of local symmetry for initial segmentation,
- flow field analysis for segmentation of moved objects,
- space-time analysis for moving object segmentation,
- color quantization for the extension of the feature base,
- use of nonlinear dynamics for lateral and longitudinal control,
- texture analysis, e.g., determination of local entropy or cooccurrence measures for initial segmentation,
- use of cross entropy measures for object tracking,
- use of order statistics for object tracking,
- parametric optimization and local adaptation,
- Hausdorff distance measures for object tracking and classification,
- cooccurrence-based measures for classification,
- fusion of different representations and sensors,
- neural networks for various classification and fusion tasks.

A few selected methods developed at the Institut für Neuroinformatik will now be presented in more detail.

1.1 Local Orientation Coding

Christian Goerick, Thomas Kalinke

The 'raw' gray scale images are preprocessed by a method we call local orientation coding (LOC). The image features obtained by this preprocessing are bit strings each representing a binary code for the directional gray-level variation in a pixel neighborhood. In a more formal fashion the operator is defined as

$$b'(n,m) = \sum_{i,j} k(i,j) \cdot u(b(n,m) - b(n+i, m+j) - t(i,j)), \quad (i,j) \in \text{ neighborhood}$$

where $b(n,m)$ denotes the (gray scale) input image, $b'(n,m)$ the output representation, $\mathbf{k}$ a coefficient matrix, $\mathbf{t}$ a threshold matrix and $u(z)$ the unit step function. The output representation consists of labels, where each label corresponds to a specific orientation of the neighborhood. For a N_4 and a N_8 neighborhood on regular square grids, suitable choices for the coefficient matrices are

$$\begin{bmatrix} 0 & 1 & 0 \\ 2 & R & 4 \\ 0 & 8 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 4 \\ 8 & R & 16 \\ 32 & 64 & 128 \end{bmatrix} \quad \begin{matrix} n & \\ \uparrow\!\!\!\hookrightarrow & m \\ & \end{matrix} \quad ,$$

where R is the reference position. This choice for N_4 leads to a set of labels $b'(n,m) \in [0, \ldots, 15]$ corresponding to certain local structures. The choice of the coefficients and the formulation of the operator gives rise to some properties:

- Due to the unique separability of the sum into its components, the information of the local orientation is preserved.
- The approach is invariant to absolute intensity values.
- The search for certain structures in the image reduces to working with different sets of labels. For horizontal structures mainly the labels 1, 8 and 9 have to be considered.

An adaption mechanism for the parameters $\mathbf{t}$ of the coding algorithm yields a high level of flexibility with respect to lighting conditions [GNW96] [Goe94].

1.2 Symmetry

Thomas Kalinke

One powerful feature is the bilateral symmetry [BG85]. It is not only found in technical and industrialized surroundings where mainly automation processes have to be realized, but especially in nature. In the process of evolution the symmetry has developed as an indicator for food and prey or enemy [SK92]. Plants and animals often show a symmetric structure, where especially in the case of animals the direction of movement mostly corresponds to the symmetry axis. The detection of symmetry features of animals which appear or flee is an important task in the question of surviving in nature. The symmetry of plants for example trees should not be viewed as the exact correspondence of the contours on both sides of their shape, but as a degree of similarity. The ability of man to detect symmetry as an object feature results from a step of high visual processing [BT90]. One tends to extract symmetric regions as closed areas even if the images are highly unstructured. The detection of vertical symmetry axis is the most reliably developed in nature [Roc84]. In [Roy81] it is shown that compared to all others the detection with the fastest reaction time is found for objects consisting of vertical symmetry axis. Mainly all algorithms for symmetry detection can be parted into intensity- and/or edge-based [Ros86]. An algorithm using even- and odd-functions of the image pixel intensity distributions for symmetry axis detection is presented in [Mar89]. The edge-based methods try to determine the symmetry axis by finding best matching features. A combination of both is presented in [ZBvS93]. Approaches using neural networks are shown by [BS89] (binary images), [Lis92] ,and [JLS94] (binary images).

Our interest in vertical symmetry detection originates from the object detection task. Especially vehicle front and back views are strongly symmetric. Furthermore, we want to detect vehicles which are situated in a long-distance area as early as possible to increase the ability of scene analysis before the object is in the nearer surrounding. Under these circumstances the system has to cope with low pixel resolution, noise, partial occlusions of objects and the signal-to-noise ratio of the background to the object in given hypotheses (region of interest) may be low. We developed a neural network that is capable of measuring the strength of symmetry. Furthermore a set of different positions of the symmetry axis are under inspection at a time. The lack of symmetry is plotted in figure 1 for different hypotheses of symmetry positions. The object covers only half of the image.

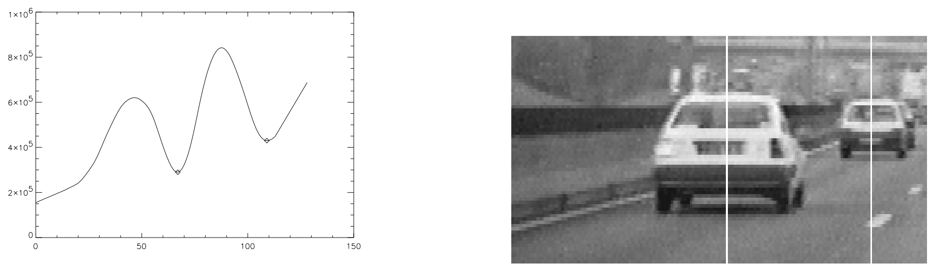

Figure 1: Plot of the lack of symmetry over possible horizontal positions in the image. Detected minima correlate with maxima of symmetry strength.

1.3 Flow Field Analysis for Motion Segmentation

Gesa Lorenz, Martin Werner, Walter Gillner

On motorways, vehicles overtaking or driving into lanes potentially increase danger for all road users. Thus their detection is a vital part when designing a computer guided, partly autonomous vehicle control.

In our approach the detection problem is split up into several processing steps. At first we generate object hypotheses by using a fast algorithm for motion segmentation. These hypotheses are either accepted or rejected by classification and tracking procedures which are already integrated into our overall system.

The search for initial regions is limited to the right and left part of the image where the objects of interest are expected. Because of this the ego- and object-motion can be postulated to have only translatory components. Later, when analyzing the motion components, this will prove to be advantageous. The size of the search space is adapted to the camera geometry as well as to the lane geometry.

The motion segmentation is carried out by a feature based optical flow approach permitting fast computation. It consists of matching corresponding image features extracted from two successive frames.

Figure 2 (a) shows a flow field typical for an overtaking manoeuvre on a motorway. Due to the ego motion of the observer, flow field components arise which diverge radially from the focus of expansion and consequently have different horizontal directions in the side parts of the image. Independent object motion represented here by the overtaking vehicle creates a deviating flow field that can be used for object detection.

The data are preprocessed by *local orientation coding* (LOC), an approach which has been developed at our institute [Goe94]. As a result of this preprocessing step we obtain features encoded as bit-strings which represent the direction of gray-value differences in a neighborhood of 4 pixels. In order to compute the flow field we find correspondences between the previously extracted features of two successive frames. More precisely we search in a neighborhood of 8 pixels for a unique and perfect match. To increase resolution we calculate on an image pyramid. Thus larger displacements and therefore larger velocities can be detected.

The following figure 2 (b) shows the gray-value image, the LOC image, as well as the flow field. The flow field is divided into two different images according to its horizontal direction. Having a component pointing to the right it is plotted in the left image, otherwise in the right one. The flow field created by the overtaking vehicle appears in the left part of the left image. Due to its position it can easily be separated from the flow field induced by the ego motion. This is done by computing and thresholding the histograms of the flow field along the image coordinates. The result is represented in the original image by a white frame. A temporal prediction stabilizes our algorithm.

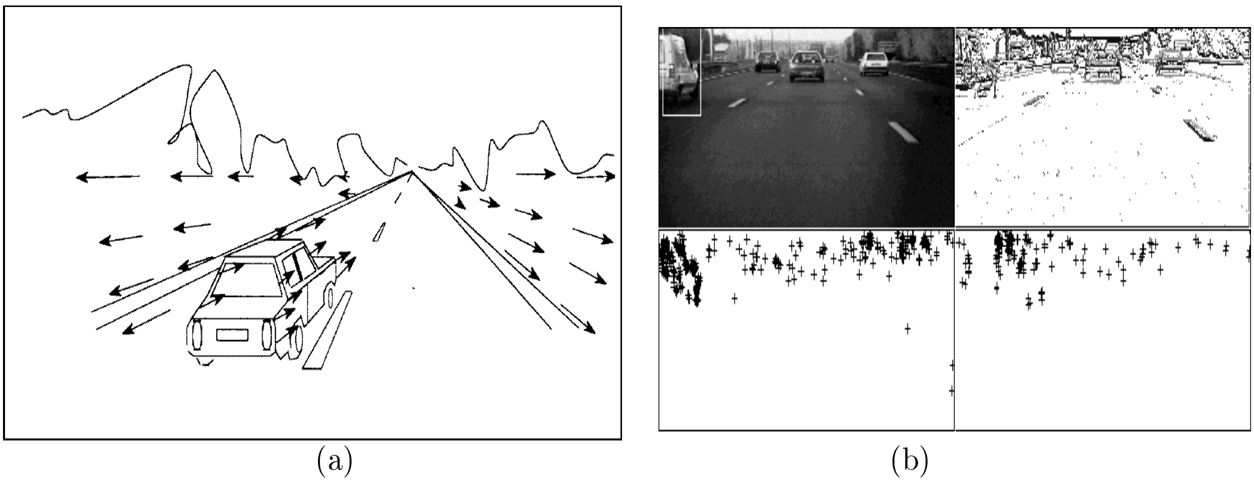

Figure 2: (a): Flow field. (b): from top left to bottom right: gray-value image, LOC-image, flow field with components to the right, flow field with components to the left

1.4 Space-Time Analysis

Iris Leefken, Walter Gillner

A moving observer given the task of detecting moving objects is confronted with the problem of differentiation between ego-motion and object-motion.

Two concepts using the time dependency of image sequences were developed to achieve segmentation of the images into moving objects and background. For this the effect of ego-motion has to be eliminated. The first architecture is based on optical flow as described in section 1.3. The second one makes use of space-time analysis.

In space-time analysis, the temporal variation of gray values sampled from images is evaluated. The gray values are sampled along trajectories determined by the predicted course of ego-motion. Space-time diagrams can be constructed from this data. These diagrams can be used for determining the speed of objects moving along the trajectories. The speed is measured relative to the moving observer (figure 3). Given the speed of the objects, the image can be segmented into regions of different object speeds. If the speed of the observer is known, the effect of ego-motion can be eliminated in the resulting image. This can be achieved by consideration of regions in which the speed differs from the speed of the observer (figure 4). The distinction between moving objects can be achieved by taking into regard either the different speeds in separate speed regions or the different location of the speed regions.

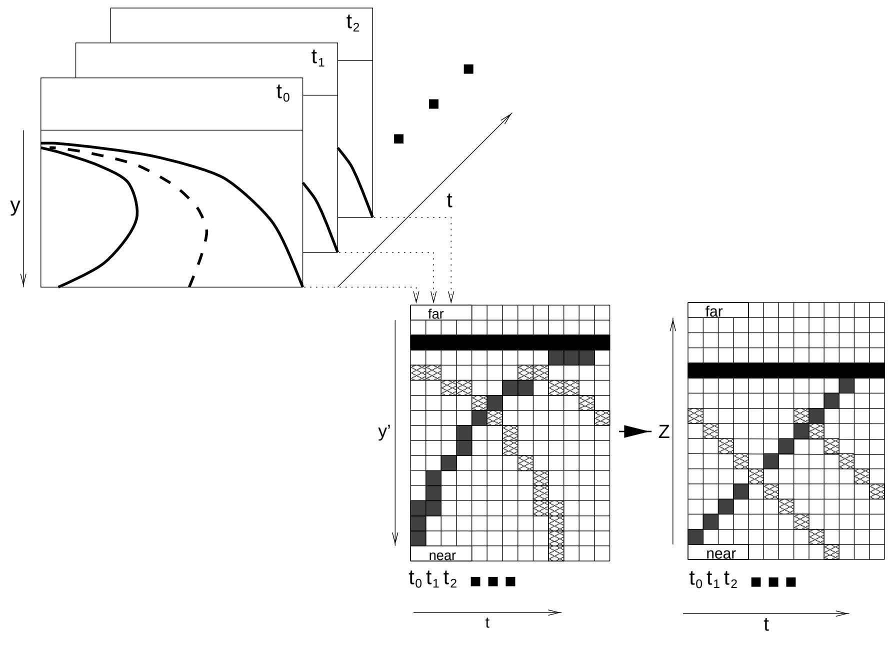

Figure 3: Sampling of gray-values along a trajectory (dashed line). Projection of the space-coordinate along the trajectory onto the vertical axis of the space-time diagram. By back-projection the measure of the axis along the sampling trajectory (y'-axis) can be transformed to the distance to the objects in the plane of movement (Z-coordinate).

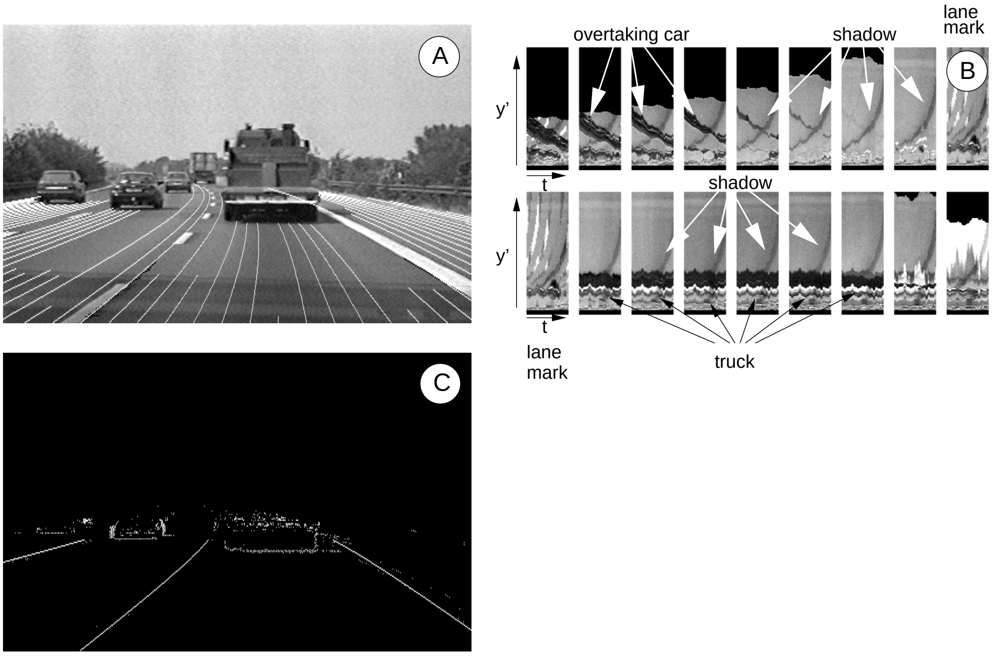

Figure 4: Segmentation of a highway scene. The original image containing the sampling trajectories is presented in figure A. Figure B shows the space-time diagram for all trajectories. In Figure C the corresponding binary image after elimination of ego-motion is displayed. The traffic lanes are marked for better orientation.

1.5 Color Quantization

Gesa Lorenz

In image analysis tasks, color information offers a valuable supplement to the commonly used gray-value information. A significant measure in that context is color contrast, also of primary importance in the human visual system. By using color information, a distinct gain of stability and robustness is obtained.

Different advantages are provided by integrating color information in the feature space of our overall system. On the one hand, color as a surface property of objects enables a transition from edge-based to area-based object descriptions. On the other hand, color information itself strongly influences the attention of the human observer, especially in the context of traffic scenes. Therefore it can also be used for automated attention control.

For generating areas of constant color, we applied a color quantization method based on the *neural gas* algorithm [MBS93]. The underlying structure is a self-organizing neural network which arranges a selectable number of codebook vectors in color space approximating the color distribution of the image pixels. By replacing the original color values with their nearest codebook vector we gain image regions with homogenous color values, i.e. color areas. The implemented neural gas algorithm is distinguished from other neural networks by small reconstruction error and high velocity of convergence. Furthermore, after a unique initialization of the codebook vectors, it can adapt to changing image statistics with a relative small effort so that image sequences can be handled. The result of a color quantization using 16 codebook vectors can be seen in figure 5. When regarding the vehicles or the sky, regions of homogenous colors can be observed.

The described area coding is presently employed by two object tracking algorithms using Hausdorff distance (section 1.8) and cross entropy (section 1.6.4) for object matching. These metrics utilize the distribution of color values and the edges distracted from color areas, respectively. In comparison with gray-value based techniques, stability and robustness improved. Current research work focuses on further improvement of the color quantization in order to obtain more homogenous color areas. To receive better results space information in addition to color information has to be considered.

Figure 5: Left: original image, Right: color quantized image

1.6 Texture

Thomas Kalinke

Besides operators like intensity derivation (gradients and the LOC) texture analysis as an integrating operator has been used successfully in image processing. The term *texture* is not explicitly defined. Globally texture is a description of image pixels or texture elements (groups of pixels) belonging to a specific texture class due to their spatial arrangement to other elements. Texture depends inherently on scaling. The spatial and intensity relationship between these elements define the kind of texture. Strong variation of intensity in a small area lead to fine textures and low variations produce coarse textures. Furthermore textures can be parted into properties weak and strong. Weak textures are described mostly by statistical methods. In strong textures the spatial interaction of elements are somewhat regular. Their recognition is usually accompanied by an exact definition of texture primitives (grammars). Actually two different methods for analyzing are commonly used: statistical and syntactic. In our applications we mainly work with statistical texture description. Every kind of texture is represented by a multi-dimensional feature vector in order to evaluate a statistical pattern recognition for every texture class based on suitable decision rules.

1.6.1 Local Image Entropy

The *local image entropy (LIE)* has been developed at our institute [KvS96]. In this method an estimation of the information contents of a pixel and its neighborhood is given. A saliency map is calculated so that a separation of objects and background can be evaluated. Figure 6 shows results of an initial segmentation process based on the LIE. The areas of objects and background are cut out by thresholding. A detection of road-users and the free driving space can be easily done.

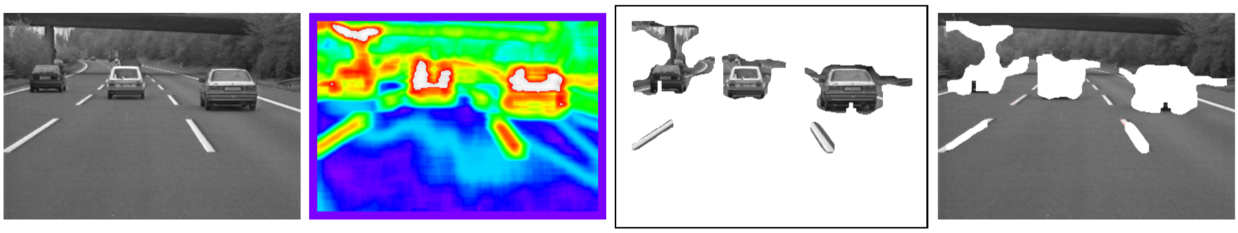

Figure 6: Image, local image entropy, image segmentation: objects and free driving space

1.6.2 Cooccurrence Matrix

One of the fundamental tools in texture analysis, the cooccurrence matrices, were suggested by [HSD73]. In here, the probability of the cooccurrences of pixel pairs under predefined geometrical and intensity constraints are measured. These constraints are determined by the intensity ratio and the spatial relationship (angle and distance) of two image points. A definition of the cooccurrence matrix follows. In an image window $\mathbf{I}$ of size $M \times N$ and a maximum number of different gray values Q the cooccurrence matrix $\mathbf{P}$ is calculated under parameters angle α in a given distance d as follows

$$P_{d,\alpha}(i,j) = \frac{\text{Number of pairs } {\scriptstyle (x,y),(x',y')}\text{, verifying } {\scriptstyle (d,\alpha)} \text{ and } {\scriptstyle I(x,y)=i} \text{ and } {\scriptstyle I(x',y')=j}}{\text{Number of all pairs in image window } {\scriptstyle (x,y),(x',y')}}$$

A calculation of texture features is performed in most of the applications under four directions ($\alpha = 0, 45, 90,$ and 135) and different distances $d = 1, 2, \ldots$. A rotation-invariance can be obtained by accumulation of the matrices of the four directions. The amount of scaling variance can be reduced by calculating the matrices over different distances.

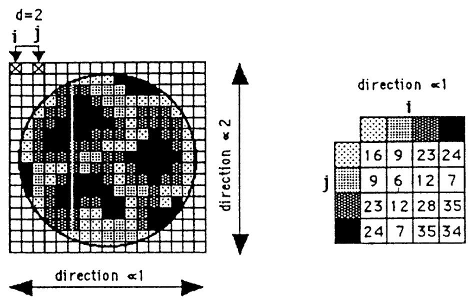

Figure 7: Image window containing 4 different intensity values and the correlating cooccurrence matrix for $\alpha = 0$ (horizontal direction) and $d = 2$.

Julesz showed that the human perception of texture is based on cooccurrence statistics. Haralick, Shunmugan, and

Dinstein suggested in [HSD73] 14 different statistical features which can be obtained from the cooccurrence matrices.

In our field of research cooccurrence matrices are mainly applied to the initial segmentation. The matrices are calculated in overlapping windows. Features like energy, entropy, contrast, correlation and the highest cooccurrence of [HSD73] are combined for the segmentation process.

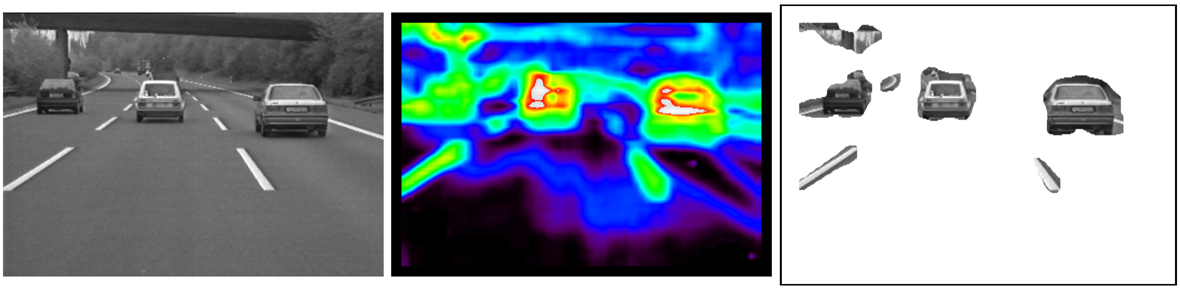

Figure 8: Image, Cooccurrence based saliency map and the thresholded and re-mapped segmentation result

1.6.3 Fractal Dimension

Another approach of texture description is based on the fractal dimension. In [Pen84] the interdependence between the fractal dimension and the structure of texture is shown. A class of texture measurements based on the fractal dimension was introduced by [KCC89]. A lane detection using the "Hausdorff-Besicovitch"-dimension [PNHA84] and a neural network has been realized in [YdSB93]. Our goal is the initial segmentation based on this measurement. Figure 9 shows segmentation results for a typical traffic scene.

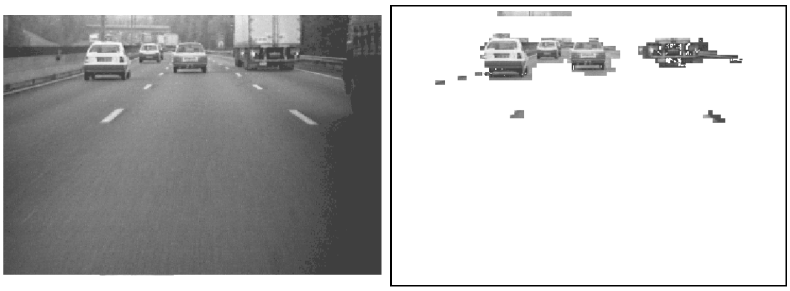

Figure 9: Intensity image, segmentation result based on the local fractal dimension analysis.

1.6.4 Texture based Object Tracking - Cross Entropy

One of the simple description of textures is obtained by intensity histograms (first order statistics). Especially non-rigid objects like pedestrians and two-wheeled vehicles which consist of a further rotational degree of freedom compared to other road-users can be tracked using the cross entropy. As described in [KvS97] a matching process can be performed by comparison of two probability distributions. In our application a model distribution at time step $(t-1)$ is compared to several hypotheses at time t. Figure 10 and 11 show tracking of pedestrians using intensity and edge probability distributions.

Figure 10: Tracking of pedestrians based on the cross entropy based on intensity distributions.

Figure 11: Tracking of pedestrians based on the cross entropy of LOC features.

1.7 Estimation of Location and Scale Parameters

Martin Werner

In computer vision applications often the need for tracking an object prior to its classification arises. In this case no model based approach can be applied. In order to cope with this demand we apply a model free approach to object tracking. It is based on the estimation of the location and scale parameter of a feature distribution by using asymptotic properties of order statistics. The basic distribution considered is the distribution of the coordinates of some kind of low level localized features, e.g. edges.

The key point of this approach is the estimation of the location and scale parameters on the basis of the inverse cumulative density function (icdf) of the signal.

The coordinates of the features are interpreted as a realization of a two dimensional distribution. In order to be able to deal with order statistics we concentrate on the marginal density in x-direction. The result easiliy extends to any other projection.

It can be shown, that in some interval the inverse cdf of the features F_n^{-1} at time t_n is related to the icdf F_0^{-1} at time t_0 by

$$F_n^{-1}(x) = \theta_1 + \theta_2 F_0^{-1}(x)(c_1' x + c_2') \quad ,$$

where θ_1 and θ_2 are the location and scale parameters, respectively. c_1' und c_2' refer to the background. The actual estimation of the icdfs is based on the asymptotic properties of the feature's order statistics [ABN92][Dav80]. The parameter estimation is carried out by optimizing a quadratic error function.

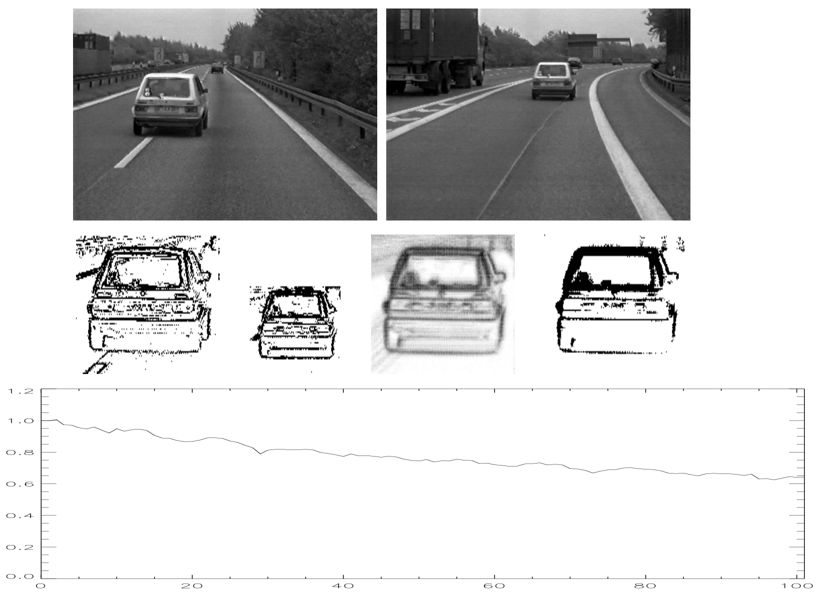

Figure 12: Frame 1 of a tracking sequence, frame 100 of the tracking sequence ($\Delta t = 50ms$), features of object in frame 1, features of object in frame 100, accumulation result of rescaled and retranslated features, a binary image obtained by thresholding the accumulation, result of the scale estimation (top to bottom, left to right.)

1.8 Hausdorff Distance for Object Tracking and Classification

Christos Tzomakas, Martin Werner, Thomas Kalinke

The geometric comparison of shapes is a fundamental tool for model-based object recognition. Most of the methods used in object recognition systems refer to a similarity measure between the model features and the image features [Pag91]. The *Hausdorff distance* measures the divergence of a set of features with respect to a reference set of features [Hut93]. These sets mostly describe object contours in our application. The comparison of similar object contours yields small distance values, whereby objects with different contours yield larger distances.

The directed Hausdorff distance h of one point set A against a point set B is the maximum of the minimum distances of each point of set A to the points of set B. The final Hausdorff distance H is simply the maximum of the two directed distances.

$$\begin{aligned} h(A,B) &= \max_{p \in A} \min_{q \in B} \|p - q\| \\ H(A,B) &= \max(h(A,B), h(B,A)) \end{aligned}$$

The partial Hausdorff distance performs a ranking of these minimum distances and considers a fraction of them instead of the maximum.

Unlike the classical correlation methods the Hausdorff distance uses Min-Max operations instead of multiplications, so it is more efficient in time. The partial Hausdorff distance is robust against partially occluded objects and outliers that may arise at the contours due to noise or insufficient feature extraction.

The partial Hausdorff distance can examine object hypotheses in a complex scene. This method was tested successfully with highway-traffic scenes. It was able to recognize vehicles on highways and track them over time. Two degrees of freedom were considered in our schema: translation and scaling of models.

Furthermore, the geometric property of the Hausdorff distance leads to the idea of classifying various vehicles into separate classes according to the imposed dissimilarity measure. Because of the need of defining a reference contour for each class we deal here with a model-based approach. The design of accurate models (prototypes) is of great importance for our task.

Figure 13: Each region is compared with two models

At a first step, the Hausdorff distance is used for the classification of cars and trucks. Due to the fact that rear views of cars differ significantly from rear views of trucks, one can expect that the design of generic models for each class can accomplish the separation of the objects of both classes. The classification works according to the following scheme: Each region is compared with two models, i.e. a car model and a truck model. The features of the region and the models have been extracted using the Local Orientation Coding (section 1.1, [Goe94]). For robuster results the horizontal features are separated from the verticals, for both the region and the models.

The Hausdorff distance is computed for each model over all the possible translations inside the region and a certain range of scales. The fractions of the features of the forward and the backward match that verify a given distance threshold constitute for each model the criteria for its classification. These values are learned by a multilayer perceptron (MLP) network using the back-propagation algorithm.

1.9 Parametric Optimization

Uwe Handmann, Detlev Noll, Martin Werner

In this section a feature based approach to object recognition and parametric optimization is described that is closely related to the elastic net approximation [DW87], resembling its advantageous optimization properties. The correspondence problem is solved by optimization of an energy function with a deterministic annealing approach.

In deterministic annealing approaches the annealing parameter influences the shape of the energy function. In the beginning the energy landscape is convex and the global optimum can easily be obtained. By decreasing the annealing parameter this optimum moves and more and more local optima arise. However, the global optimum can be tracked by local optimization approaches. An analysis shows [Nol96] that the energy function F defined by

$$F = -\frac{1}{\beta}\sum_j \ln \sum_i s_{ij} e^{-\beta d_{ij}^2} + \frac{\alpha}{\beta}\sum_j \left|\hat{\vec{y}}_j - \vec{y}_j\right|^2$$

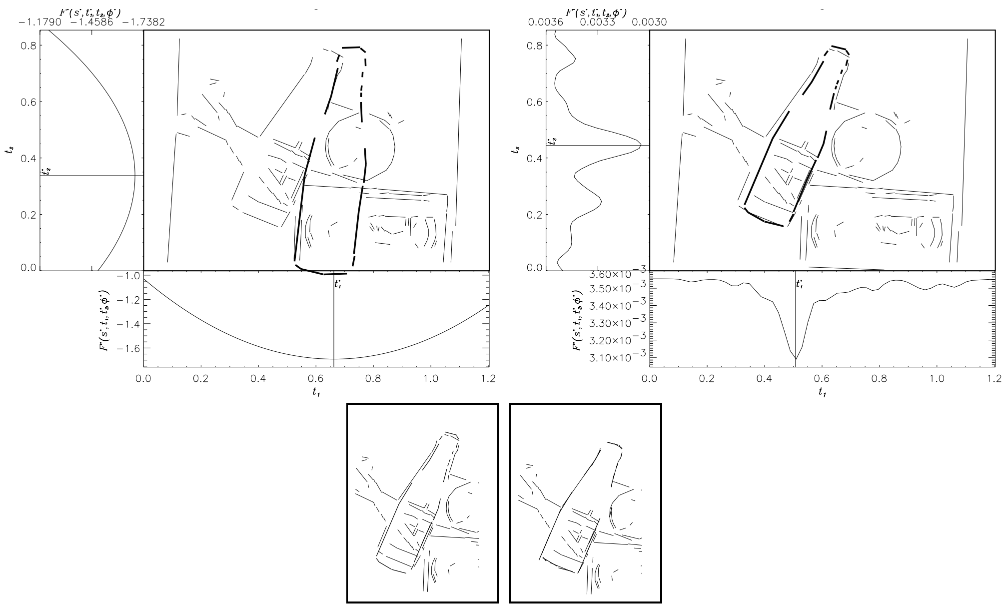

Figure 14: Top: The model (bold) overlayed the image features for $\beta = 1$ (left) and for $\beta = 1540$ (right). In addition, one dimensional cuts through F are shown. Bottom: The model after global optimization (left) and after local optimization (right).

exhibits the desired behavior of a deterministic annealing function under variation of the annealing parameter β. The parameters s_{ij} describe the feature similarity, d_{ij} describe the feature distance and α is a weighting parameter. A measurement for the local deformation can be defined by $E(\{\vec{y}_j\}) = \sum_j |\vec{y}_j - \hat{\vec{y}}_j|^2$, where $\{\vec{y}_j\}$ are the positions of the model features and $\{\hat{\vec{y}}_j\}$ is a reference model.

A first example of the algorithm's behavior is shown in figure 14. Straight line segments are the basic features used in this example. A model of a bottle (bold lines) is mapped on the features extracted from an image. The model is allowed to scale, translate and rotate in the plane. In the first row the position of the model at small and large values of β is displayed, respectively, along with one dimensional cuts through the four dimensional energy function F. The second row shows the result after the local optimization (right) in comparison to the results after the global optimization (left). An extension of the algorithm to 3D object recognition is straight forward [Nol96].

In the example of figure 14, β has been exponentially increased. The initial value depends on the a-priori knowledge. Larger values can be used, if an initial estimate is given, e.g. in object tracking applications. The final value of β depends on the desired degree of accuracy. Even if the functionality of our approach is independent of a particular choice of the local optimization algorithm, the rate of convergence strongly depends on this choice. For example, using an expectation-maximization algorithm for local optimization [Nol96, YSU94], an implementation of this algorithm for the detection and tracking of the rear view of vehicles in traffic scenes (figure 15) typically takes 50-100ms on a SUN Sparc 20.

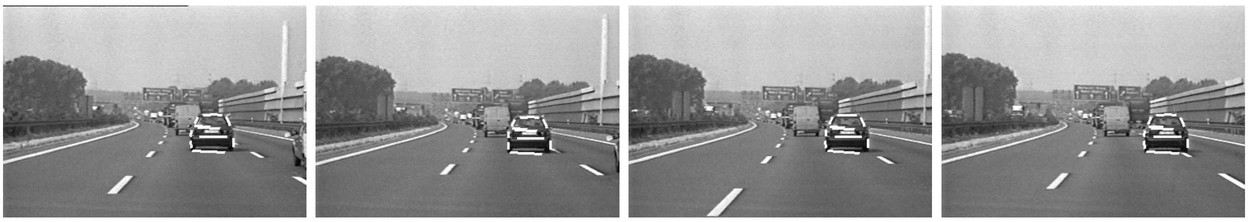

Figure 15: Tracking sequence (shown with a skip of 10) with a model of the rear view of a car (white)

1.10 Fusion for Initial Segmentation

Uwe Handmann

The aim of fusion in computer vision is to get an improvement of special solutions and single methods with a coupling net. Especially the modular coupling of single processing steps generates redundancy necessary for object recognition. Within this, greater flexibility and robustness of the image processing modules and higher adaptation of the modules regarding to the problems should be achieved.

1.10.1 Neural Coupling

A formulation of a robust and error tolerable segmentation is shown in figure 16. Computer vision modules, generating lines (polygon approximation of the contour), entropy [KvS96], and local orientation coding (LOC,[Goe94] [GNW96]) are coupled in a neural network (figure 16). A feedback over time is possible, additional sensor information could be easily integrated at this level.

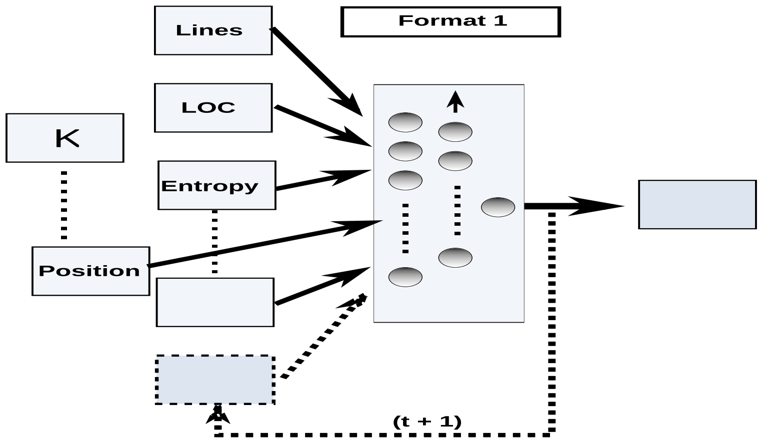

Figure 16: coupling model

1.10.2 Segmentation by Fusion

Based on integrative (entropy) and differential (LOC, lines) representations a 12-dimensional input vector

$$\mathbf{u}(x,y)^T = (\mathbf{u_1}(x,y)^T, x, y)^T$$

for the coupling net is generated for each pixel. Here, the 10-dimensional vector $\mathbf{u_1}(x,y)$ refers to the contour and texture information and the values of x and y represent the pixel's coordinate position. Vector $\mathbf{u_1}(x,y)$ is defined by

$$\mathbf{u_1}(x,y) = \sum_{(i,j)\in R} \mathbf{v}(i,j)$$

where R is a local neighborhood (e.g. 9×9) of (x,y) and $\mathbf{v}(i,j)$ is a binary vector. $(v_1(x,y),\ldots,v_4(x,y))^T$ code a subset of the LOC results, $(v_5(x,y),\ldots,v_9(x,y))^T$ code the value of the entropy and $v_{10}(x,y)=1$, if (x,y) belongs to a polygon.

A multilayer perceptron (MLP, [Hay94], page 138ff) with a 12-5-1-structure of the layers is used here as a coupling net. The feature vector is propagated through the twelve input neurons and the five hidden neurons to the output. The activity of the output neuron o describes the membership of each pixel to the initial segments, if $o > 0.5$.

The back-propagation algorithm ([Hay94], page 142ff) is used for training the net. With a handlabeled database the error function between the received output activity after passing the net and the required output is calculated and propagated backwards. This structure is flexible to be enlarged and adaptable for new problems.

Figure 17 shows the robust segmentation results of cars in a sequence of traffic scenes. Only every 50th frame is shown in the figure.

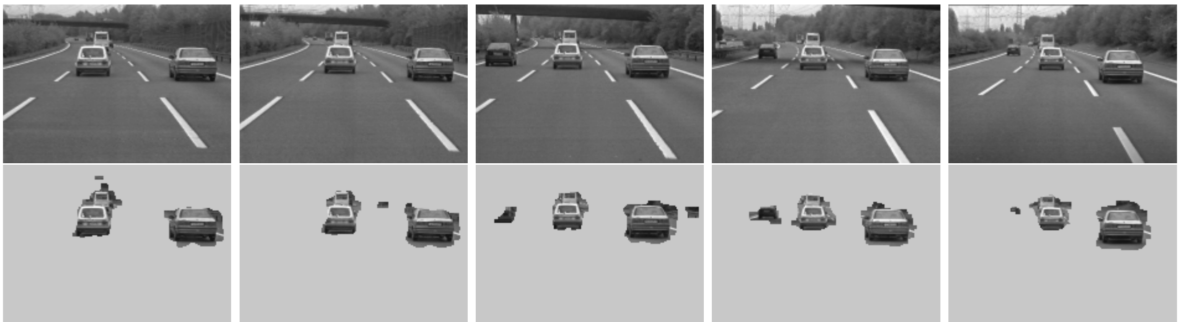

Figure 17: Initial segmentation with a MLP coupling net

1.11 LOC-Classification

Christian Goerick, Uwe Handmann, Thomas Kalinke, Martin Werner

With the given local orientation coding (LOC, [Goe94] [GNW96]), described in section 1.1), a classification of vehicles is realized. The classifier has to cope with partial occlusions,

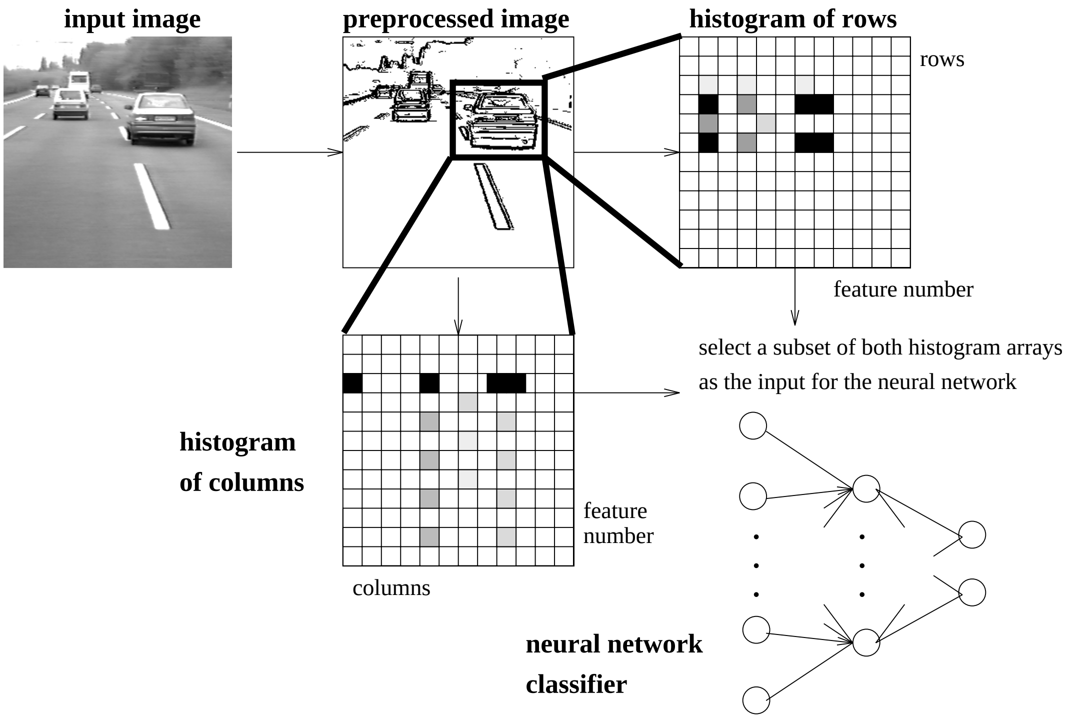

Figure 18: Schema of the detection process.

varying illumination conditions, tilt of an object, differently resolved structures depending on the distance of the object under consideration, noise and perturbations induced by the recording and processing equipment, different viewpoints and different kind of cars with different shapes and colors. Additionally, the classifier should be able to generalize from relative few training examples to the necessary features characterizing a car. Therefore, a neural network has been chosen for solving the classification task. It is a feed-forward neural network with one hidden layer trained by the error back-propagation algorithm [HKP91, RMG86].

These networks are known to be universal approximators for any continuous valued function [HSW89]. Furthermore, it is shown that these structures can, with some small modifications, approximate a-posteriori probabilities in the sense of a Bayesian classifier [FM93].

The inputs for the classifier are certain subsets of the histograms. The output is the class of the region.
The complete system has been implemented and extensively tested on the Mercedes Benz VITA II test vehicle [BGGZ94]. Different classes of vehicles have been trained. For a further evaluation of the system refer to [BZF95].

1.12 A modular Neural Classifier for Vehicles

Iris Leefken, Thomas Kalinke

Besides the task of object detection and tracking over time the process of classification is a valuable information. The large amount of different vehicle types and the interfering environmental conditions like weather conditions (shading, illumination, etc.) require high flexibility and robustness of the classifier. A-priori knowledge can be incorporated by decomposition of the classification task to subtasks. A modular neural network has been designed for vehicle classification in order to require a small input space for every module and to facilitate the analysis of the classifiers' performance. Every module is an expert in its small band application spectrum and resulting in high performance by gaining from a suitable coupling structure. The LOC- and the Hausdorff-classifier are efficiently combined

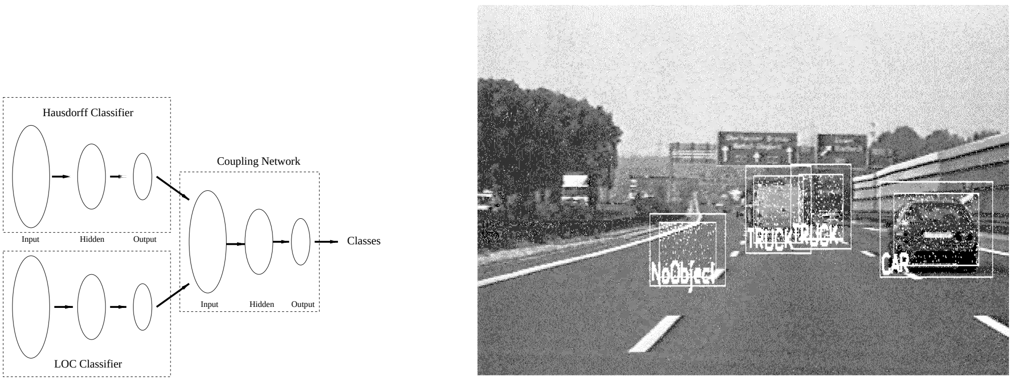

Figure 19: Modular neural network and classification results

by a modular neural net using a multilayer perceptron as a coupling network. It has been trained to distinguish cars and trucks from the background. Figure 19 shows classification results (car, truck and no object) for a typical traffic scene.

2 Detection of Detonators in Luggage

David Kastrup, Bernd Völpel

Task at hand was supplementing methods for X-ray based detection of plastic explosives in luggage examined at airports by recognition of the typically employed detonators. The X-ray systems in use are produced by the project partner Heimann Systems. Human operators working at the X-ray systems are to be assisted by region coloring of areas suspected of containing plastic explosives or their detonators. The necessary computational power is provided by parallel processing systems with few processors. It has been the task of the Institut für Neuroinformatik to develop the software for detecting the detonators.

Early development stages concentrated on the search for suitable detectable features, which was done in the simulation language IDL. A preliminary detection can be accomplished by thresholding operations combined with the application of morphological criteria (size of connected regions, characteristic local extrema). The contours are marked using LOCs (Local Orientation Codes, section 1.1), so that their characteristic outlines can be detected by determining statistics over finely grained scaled and rotated masks. The inside of a detonator is subsequently compared in a matching process with fixed luminosity distributions.

The second stage of development concentrated on porting the algorithms to the language "C" and finding optimizations for satisfying the real-time constraints of the system. Apart from the development of fast variants of various employed algorithms (i.e., for fast region labeling and fast rotation of bitmaps), several efficient lookup structures were developed for time-critical components, tested, and optimized based on profiling data. The runtime-oriented proper nesting of image processing operations turned out to be another important asset to achieving good performance. Preliminary detection of such features in the inside of potential detonators as were detectable by largely rotation and scaling invariant filter operations turned out to significantly aid in the cutting down of the number of matching operations for outline orientation and scale detection. As an added advantage the false alarm rate could be lowered without noticeable impact on the detection rate.

The project partner Heimann Systems is planning to offer the completed system, given suitable interest, as an addition to their range of security products.

3 Ocean Environment Surveillance

Carsten Winkel

Marine pollution has serious ecological consequences even for the people's health. To support cleaning operations it is necessary to efficiently detect as well as to quantify type and volume of the discharged substances. That is the only way to identify the polluter and to avoid severe problems for the marine environment. Therefore the surface of the North Sea and the Baltic Sea are observed by regular airborne surveillance flights using various sensors [instructed by the German Ministry of Transport (BMV)]. In addition to the standard sensor equipment like SLAR (side–looking airborne radar) and UV/IR scanner and cameras new sensors have been developed, supported by the German Ministry of Education, Science, Research and Technology (BMBF). These instruments, an imaging laser fluorosensor system (LFS) [GHWW96] and a three frequency band microwave radiometer (MWR) [Grü96], allow a much better analysis of pollution found at sea. As an example fig. 20 shows the MWR and UV images of an oil spill. Type, quantity and dimensions of the pollution have to be determined by a combined analysis of these various multispectral sensor data.

The comparison of the principal components of all the fluorescence spectra at 12 detection wavelengths with those of different oil types and other substances stored in a database allows the identification and classification of the type, e. g. of crude oil. A neural net is used to improve the classification and the referring estimate of the material parameters, especially the dielectric constant and to enlarge the applicability to changing weather conditions.

Due to different skin depths of the used wavelengths three microwave channels of the radiometer give information about the various dephts of the oil layers. To estimate a thickness distribution of the oil pollution it is not useful to take only the average, because every channel has got its own radiometric and geometric resolution. This is taken into account by equalizing these resolutions using the Maximum Entropy Method [Mei88] on the two images measured with the lower frequencies under the consideration of diverse constraints and a–priori knowledge. Classical supervised neural network methods cannot be used because the target values, i. e. the depths of the oil film, are not directly measurable. Knowledge about the total oil volume only allows the application of an integral form of error. With a linear method based on SVD (singular value decomposition) parameters of a model describing the dependency of measured brightness to oil thickness are to be determined. Our further aim is the development of a completely data driven method which permits to extract those correlations under different external conditions like wind or thermal radiation of the sky. A formulation of Wei et al. [WH96] seems to be useful. Results of data preprocessing and analysis show that it is very important to get reliable calibrated data. The best way to enhance the quality of the microwave measurements is to make the measurement system self-calibrating. An extensive database of multispectral environmental data which should be used for further evaluations and coupling with diverse information is still being composed.

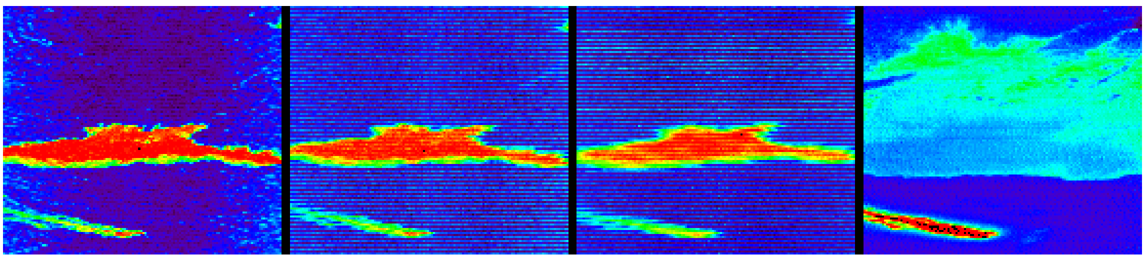

Figure 20: MWR images (taken at 89, 36, and 18 GHz, from left to right) of an oil spill in comparison to UV signatures

Bibliography

[ABN92] B.C. Arnold, N. Balakrishnan, and H.N. Nagaraja. *A First Course in Order Statistics*. John Wiley & Sons, 1992.

[BG85] V. Bruce and P.R. Green. *Visual Perception Physiology, Psychology and Ecology*. Lawrence Erlbaum Associates, London, 1985.

[BGGZ94] M.E. Brauckmann, C. Goerick, J. Groß, and T. Zielke. Towards all around automatic visual obstacle sensing for cars. In *Proceedings of the Intelligent Vehicles '94 Symposium, Paris, France*, pages 79–84, 1994.

[BS89] M. Bichsel and P. Seitz. Minimum class entropy: A maximum information approach to layered networks. *Neural Networks*, 2:133–141, 1989.

[BT90] A. Blake and T. Troscianko. *AI and the eye*. John Wiley & Sons, 1990.

[BZF95] S. Bohrer, T. Zielke, and V. Freiburg. An Integrated Obstacle Detection Framework for Intelligent Cruise Control on Motorways. In *Proceedings of the Intelligent Vehicles Symposium, Detroit*, pages 276–281, 1995.

[Dav80] H.A. David. *Order Statistics*. John Wiley & Sons, second edition, 1980.

[DW87] R. Durbin and D. Willshaw. An analogue approach to the travelling salesman problem using an elastic net approach. *Nature*, 326(6114):689–691, 1987.

[FM93] M. Finke and K.-R. Müller. Estimating A-Posteriori Probabilities Using Stochastic Network Models. In *Proceedings of the Summer School on Neural Networks, Bolder, Colorado*, pages 276–281, 1993.

[GHWW96] K. Grüner, T. Hengstermann, P. Wagner, and M.B. Wischnewsky. Recent Results Obtained from a 2nd Generation Airborne Surveillance System for Identification of Maritime Pollution. In *2nd International Airborne Remote Sensing Conference and Exhibition*, volume III, pages 791–802, Ann Arbor, Michigan, USA, 24.-27.6. 1996. ERIM.

[GNW96] C. Goerick, D. Noll, and M. Werner. Artificial Neural Networks in Real Time Car Detection and Tracking Applications. *Pattern Recognition Letters*, 1996.

[Goe94] C. Goerick. Local orientation coding and adaptive thresholding for real time early vision. Internal Report IR–INI 94–05, Institut für Neuroinformatik, Ruhr–Universität Bochum, D–44780 Bochum, Germany, June 1994.

[Grü96] K. Grüner. The Three Frequency Microwave Radiometer of a Second Generation Airborne Surveillance System for Remote Sensing of Maritime Oil Pollution. In *IEEE-Workshop 'RF and Microwave Noise'*, pages 66–69, Germany, 9.-10.6. 1996. Techn. Univ. Ilmenau.

[Hay94] S. Haykin. *Neural Networks*. Macmillan College Publishing Company, Inc., 1994.

[HKP91] J.A. Hertz, A.S. Krogh, and R.G. Palmer. *Introduction to the Theory of Neural Computation*. Addison Wesley, 1991.

[HSD73] R.M. Haralick, K. Shanmugan, and I. Dinstein. Textual features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6), 1973.

[HSW89] K. Hornik, M. Stinchcombe, and H. White. Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, 2:359–366, 1989.

[Hut93] D.P. Huttenlocher. Comparing Images Using the Hausdorff Distance. *IEEE Transactions on PAMI*, 15(9), September 1993.

[JLS94] W. Joung, C. Latimer, and C. Stevens. Modelling symmetry detection with backpropagation networks. In *Proc. of the fifth Australian Conference on Neural Networks, ACNN 1994, Brisbane*, pages 33–36, 1994.

[KCC89] J.M. Keller, S. Chen, and R.M. Crownover. Texture description and segmentation through fractal geometry. *Computer Vision, Graphics, and Image Processing*, 45(2):150–166, 1989.

[KvS96] T. Kalinke and W. von Seelen. Entropie als Maß des lokalen Informationsgehalts in Bildern zur Realisierung einer Aufmersamkeitssteuerung. In *Mustererkennung 1996*, pages 627–634, Berlin, Heidelberg, 1996. Springer-Verlag.

[KvS97] T. Kalinke and W. von Seelen. Kullback-Libler Distanz als Maß zur Erkennung nicht rigider Objekte. In *Mustererkennung 1997*, 1997.

[Lis92] P.J.G. Lisboa. Der Begriff der Entropie in der Wahrscheinlichkeitsrechnung. In P.J.G. Lisboa, editor, *Neural Networks: current applications*, chapter 1, pages 163–184. Chapman & Hall London, 1992.

[Mar89] G. Marola. Using symmetry for detecting and locating objects in a picture. *Computer Vision, Graphics, and Image Processing*, 2(46):179–195, 1989.

[MBS93] T.M. Martinetz, S.G. Berkovich, and K.J. Schulten. Neural-Gas Network for Vector Quantization and its Application to Time-Series Prediction. *IEEE Transactions on Neural Networks*, 4(4):558–569, 1993.

[Mei88] E.S. Meinel. Maximum-entropy image restoration: Lagrange and recursive techniques. *Journ. Opt. Soc. Am.*, 5(1):25–29, January 1988.

[Nol96] D. Noll. *Ein Optimierungsansatz zur Objekterkennung*. Number 454 in Fortschrittberichte, Reihe 10. VDI-Verlag, Düsseldorf, 1996. Dissertation, Ruhr-Universität Bochum.

[Pag91] D.W. Paglieroni. Distance Transforms: Properties and Machine Vision Applications. *CVGIP*, 54(1):56–74, January 1991.

[Pen84] A.P. Pentland. Fractal-based description of natural scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 1984.

[PNHA84] S. Peleg, J. Naor, R. Hartley, and D. Avnir. Multiple resolution texture analysis and classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 1984.

[RMG86] D. Rumelhart, J.L. McClelland, and PDP Research Group. *Parallel Distributed Processing.* MIT Press, Cambridge, Massachusetts, USA, 1986.

[Roc84] I. Rock. *Perception.* Scientific American Library New York, 1984.

[Ros86] A. Rosenfeld. Axial representation of shape. *Computer Vision, Graphics, and Image Processing*, 33:156–173, 1986.

[Roy81] F.R. Royer. Detection of symmetry. *Journal of Experimental Psychology: Human, Perception and Performance*, 7(46):1186–1210, 1981.

[SK92] E. Schulze-Krüger. *Analyse von Augenbewegungen des Menschen zur Symmetrie- und Raumwahrnehmung und Vergleich zu einem aktiven Kamerasystem.* VDI Verlag, Reihe 17 Biotechnik Nr.83, 1992.

[WH96] G.Q. Wei and G. Hirzinger. Learning Shape from Shading by a Multilayer Network. *IEEE Trans. Neural Networks*, 7(4):985–995, July 1996.

[YdSB93] G. Yahiaoui and M. de Saint Blancard. Texture-based image segmentation for road recognition with neural networks. In *Neural Networks and their Industrial & Cognitive Applications*, pages 321–328, 269-287, Rue de la Garenne, 92024 Nanterre Cedex, France, 1993. EC2.

[YSU94] A.L. Yuille, P. Stolorz, and J. Utans. Statistical Physics, Mixtures of Distributions, and the EM Algorithm. *Neural Computation*, 6:334–340, 1994.

[ZBvS93] T. Zielke, M. Brauckmann, and W. von Seelen. Intensity and edge-based for symmetry detection with an application to car-following. *Computer Vision, Graphics, and Image Processing*, 2(58):177–190, 1993.