Fusion of Different Sensors and Algorithms for Segmentation

Uwe Handmann, Gesa Lorenz, Thomas Schnitger, and Werner v. Seelen

Abstract— **In this article we present a system for coupling different base algorithms and sensors for segmentation. Three different solutions for image segmentation by fusion are described, compared and results are shown. The fusion of base algorithms with color-information and a sensor fusion process of an optical and a radar sensor including a feedback over time is realized. A feature-in decision-out fusion process is solved. For the fusion process a multi layer perceptron (MLP) with one hidden layer is used as a coupling net. The activity of the output neuron represents the membership of each pixel to an initial segment.**

Keywords— **Segmentation, Machine Vision, Data Fusion.**

I. Introduction

FULLY or partly autonomously guided vehicles, particularly for road-based traffic, impose high demands on the development of suitable algorithms. This is due to the conditions imposed by natural environments. At the Institut für Neuroinformatik in Bochum, Germany, present projects are concerned with the analysis of traffic scenes [1]. In principle the analysis of these scenes is a hierarchical process with a segmentation, a classification, and a tracking task.

The segmented picture represents a substantial aspect of the automatic scene analysis. By segmenting a partitioning of the picture in object hypotheses and background is understood. The generated object hypotheses are classified and tracked within the total system in further processing steps.

The great variety of different traffic scenarios as well as the high degree of reliability necessary for the given task require an encompassing and flexible system architecture [2]. The variety of geometric appearances of involved objects and those 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 the structure of the system has to be adaptable to allow accommodation of additional modules without degeneration of already accomplished partial solutions. For this reason, even 'simple' applications are encumbered by considerations concerning the system architecture.

In this article three solutions of the fusion process for segmentation are presented. First, a simple application is described. Second an integration of color information is shown and third a sensor fusion process is solved.

Uwe Handmann, Gesa Lorenz, Thomas Schnitger, and Werner v. Seelen are with the Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany

II. Fusion process

Essentially the total system can be divided into base algorithms and algorithms for the fusion process. The base procedures supply special partial solutions with given boundary conditions. The results of the individual algorithms are not independent, so the fusion of the results entails an increase of redundancy making the total system safe and reliable ([3], [4], page 32). The algorithms of the fusion process provide for a flexible interaction and for an integrative result of the base components.

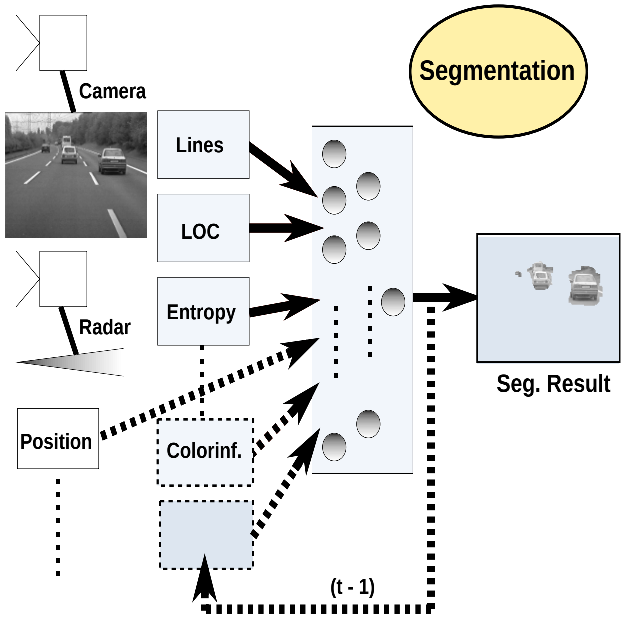

Fig. 1. Coupling model

A suggestion of a fusion process is shown in figure 1. Base algorithms, generating lines (polygon approximations of the contour), local orientation coding (LOC [5], [6]), and local image entropy [7] are coupled. As coupling structure a multi layer perceptron (MLP [8], page 138) is used. Here the coupling is learnable and flexible [3]. An on-line learning is just as possible as a feedback over time. Additional sensor information (radar, lidar, position) or other base algorithms can easily be integrated at this level. The flexibility of the coupling structure is clarified with special

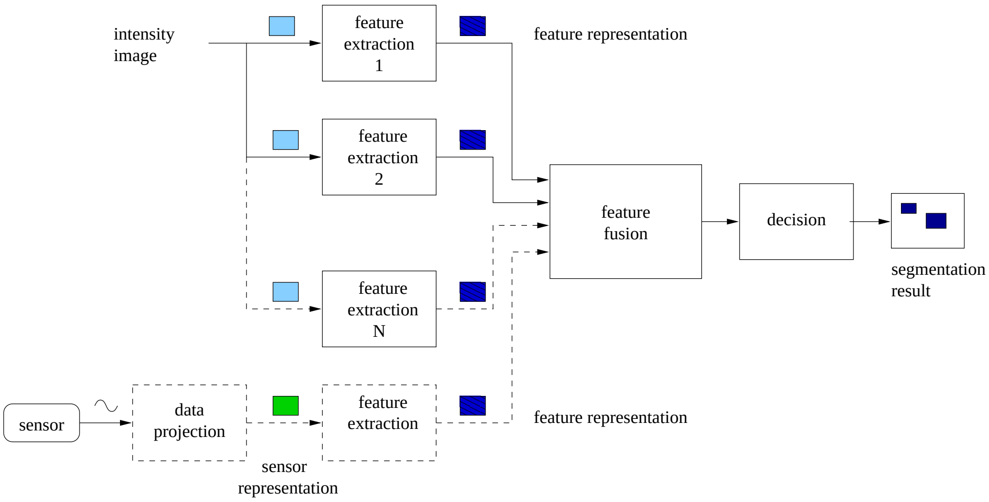

Fig. 2. Principle of a multi sensor fusion on the feature level

solutions. Three examples of the segmentation task using the fusion process are implemented.

A fusion process can take place on different hierarchical levels ([9], page 6). In general, a fusion process can be established on the data -, feature or decision level. In the present case a fusion process on the feature level is selected (figure 2), in order to use the advantage of the data reduction compared with a fusion on the data level. Relevant result features of the base algorithms are assembled to build a feature vector. The features are fused with the help of a neural net. A threshold at the output activity (decision) assigns the individual pixels of a frame to the background or a relevant segment.

III. Segmentation

In this chapter three realizations of the fusion process to solve the segmentation task are described. The results are shown on different traffic scenes with specific requirements. First a simple fusion of differentiating and integrating base algorithms is shown to clarify the principle of the fusion process on the feature level. In the second part the integration of color information into the fusion process is realized. A third implementation shows a complex feature fusion process. A sensor fusion, using optical and radar information, is realized. Additional base algorithms and a feedback over time are integrated.

A. *Segmentation based on intensity image sequences*

In this part differentiating and integrating base algorithms (figure 3) couple into a neural net, in order to solve the segmentation problem. The polygonal approximation of the contour and the LOC describe the differentiating features. The local image entropy is used as integrating feature. For each pixel a twelve-dimensional input vector

$$\mathbf{f}(x,y) = (\mathbf{f_1}(x,y)^T, x, y)^T \tag{1}$$

for the coupling net is generated. The 10-dimensional vector $\mathbf{f_1}(x,y)$ includes the contour and texture information, the variables x and y represent the pixel coordinates. The vector is defined as

$$\mathbf{f_1}(x,y) = \sum_{(i,j)\in R} \mathbf{u}(i,j). \tag{2}$$

whereby R describes the local neighborhood (9×9) of a pixel (x,y) and $\mathbf{u}(i,j)$ is a binary vector. The vector items $u_1(x,y), \ldots, u_4(x,y)$ encode reduced LOC features, $u_5(x,y), \ldots, u_9(x,y)$ encode the different entropy values and $u_{10}(x,y)$ is set, if the pixel is part of a polygon.

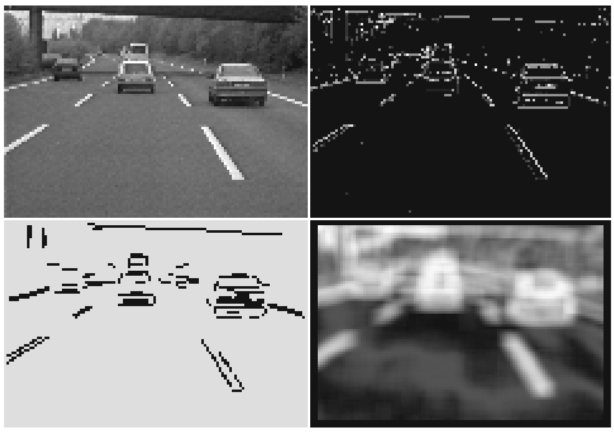

Fig. 3. Intensity, local orientation coding, polygons, local image entropy

As coupling net a MLP with a 12-5-1 structure is used. The twelve-dimensional feature vectors are propagated forward across twelve input neurons and five hidden neurons to the output neuron of the net. The activity of the output neuron represents the affiliation of the pixels to the initial segments. The decision is made by a threshold value. The training of the coupling net

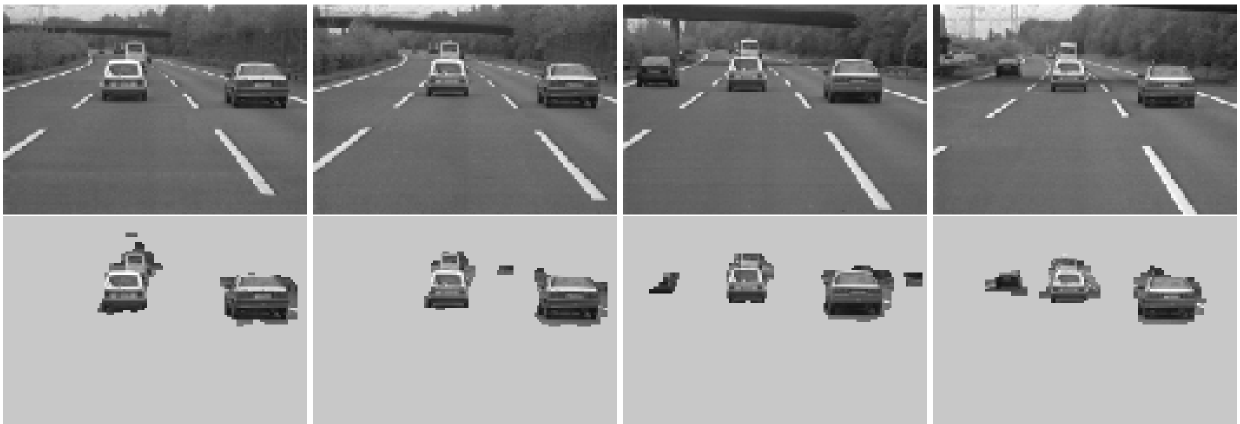

Fig. 4. Segmentation with a MLP fusion process

is performed by the back propagation algorithm ([8], page 142). The error function between output activity and the specified condition at the output (hand segmented data) is calculated and propagated backwards through the net.

Figure 4 clarifies the segmentation result of vehicles in a sequence of 200 frames of a traffic scene, whereby each 50th frame is represented. All relevant objects are segmented stably. However partial false segmentation of objects with small contrast or missing structure is possible (figure 4, column 3).

For analysing more complex scenarios a solely standard camera (CCD[1]-Camera) based approach is not sufficent. On the one hand the dynamic range of these cameras is small and on the other hand only the intensity of the scenarios is analyzed. These disadvantages can be avoided by the use of other cameras (e.g. HDRC[2], color camera) and the inclusion of further sensors.

B. Integration of color saturation for segmentation

Color is an important feature evaluated by the human visual system to solve complex tasks like scene analysis or control of attention. Both in nature and in man made environments the signalling effect of color is used to arouse attention. Thus, objects with special importance in road traffic, like traffic signs, brake lights, ambulances, etc. are marked with salient colors. Therefore, several applications in the field of visually guided driver assistance systems successfully apply color information for detecting objects with characteristic color distributions [10][11].

Dealing with more general detection tasks, color features have been used rarely so far since color is generally not a specific property of objects. Vehicles are characterized mainly by a distinct form. In image analysis this form is often represented by contours coded here as lines and LOC features. Furthermore, they possess a distinct texture in comparison to the surrounding road surface. This property is expressed by the local image entropy. Numerous vehicles are characterized by a salient color. The use of color in addition to form and texture information improves the segmentation in situations where the exclusive use of the other features often fails, e.g. the segmentation of objects with large homogeneous areas.

The fusion concept proposed here offers the possibility to integrate color information into the coupling net apart from the already available differentiating and integrating features.

Color is described adequately by three quantities. Therefore, to allow a convenient specification of colors a 3-D coordinate system has to be established. In order to avoid intensive computation often the hardware-oriented RGB color space is chosen. We are interested in extracting highly saturated image areas independent of their actual hue or intensity value. Therefore, a color description based on the HSV color space describing color by the attributes hue, saturation and brightness ([12], pp. 590) is used. For each pixel only the saturation coordinate is evaluated

$$S = \frac{\max(R,G,B) - \min(R,G,B)}{\max(R,G,B)} \quad . \tag{3}$$

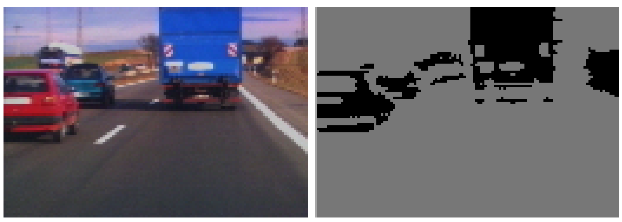

Fig. 5. original image, thresholded saturation image

Thresholding the saturation image $s(x,y)$ with s_{min} results in a binary image

$$v(x,y) = \begin{cases} 1 & \text{if } s(x,y) > s_{min} \\ 0 & \text{otherwise} \end{cases} \quad , \tag{4}$$

[1]Charged Couple-Device
[2]High Dynamic Range Camera

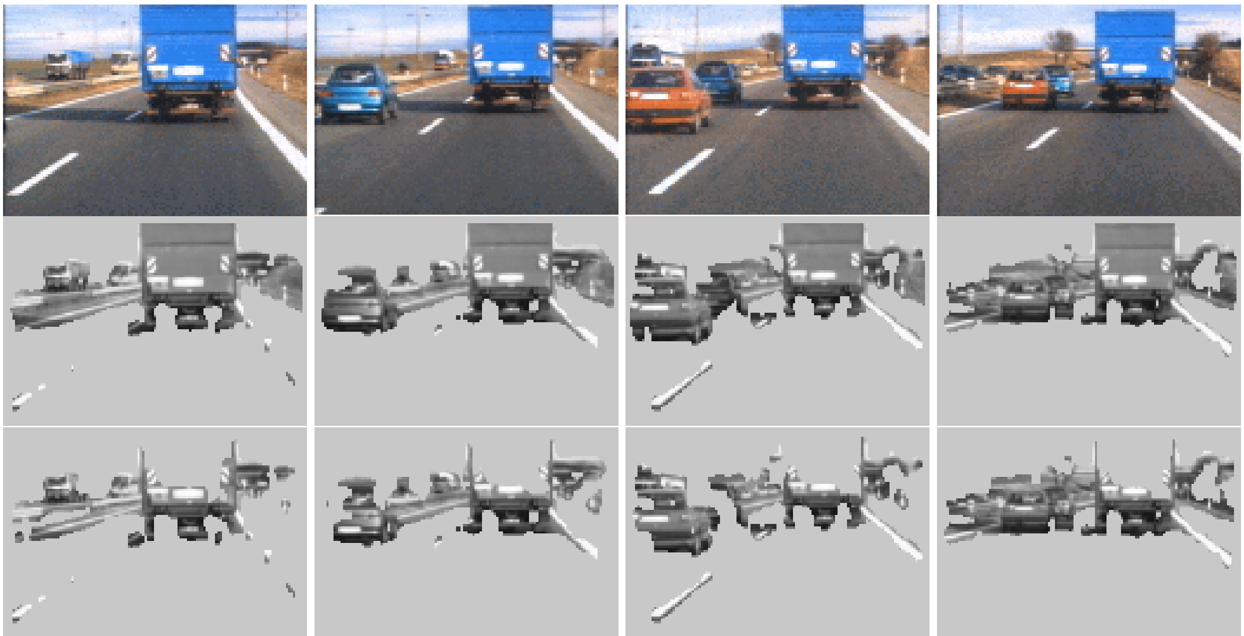

Fig. 6. Segmentation using a MLP coupling net with/without integration of saturation

serving as a basis for the new color feature

$$\mathrm{f}_2(x, y) = \sum_{(i,j) \in R} v(i, j). \quad (5)$$

For each pixel a 13-dimensional vector is created

$$\mathbf{f}(x, y) = (\mathbf{f_1}(x, y)^T, \mathrm{f}_2(x, y), x, y)^T \quad (6)$$

and fed into the coupling net (MLP with a 13-5-1 structure). The selection of the threshold value was proved to be uncritical for the segmentation result and could be kept constant for different image sequences.

Figure 5 shows the segmentation result after thresholding the saturation image. As expected, most of the segmented regions represent the homogeneous parts of vehicles (e.g. the bodywork). Parts of the background like the vegetation at the road border are incorrectly detected though. To overcome this obstacle information obtained by additional basic algorithms or alternative sensors (e.g. radar (section III-C), lidar- or infrared sensors) are used as input for the coupling net. Furthermore knowledge of road boundaries may be used when available.

Figure 6 shows a comparison of the segmentation results with and without using the color feature in a sequence of 200 images of a traffic scenario presenting every 50th image. Whereas both methods succeed in segmenting the cars with good contrast, the additional color feature improves the segmentation results in the case of badly illuminated areas (figure 6, column 3) as well as the segmentation of trucks with large homogeneous areas (figure 6, row 3).

C. Radar and optical sensor fusion

A third complex implementation of a robust image segmentation uses besides the already discussed base algorithms a feedback over time, a local variance analysis, a shadow detection algorithm, and additional radar information to realize a sensor fusion process (figure 7). The fusion combines data from independent sensors (optical and radar) to derive information that would be unavailable from the individual sensors.

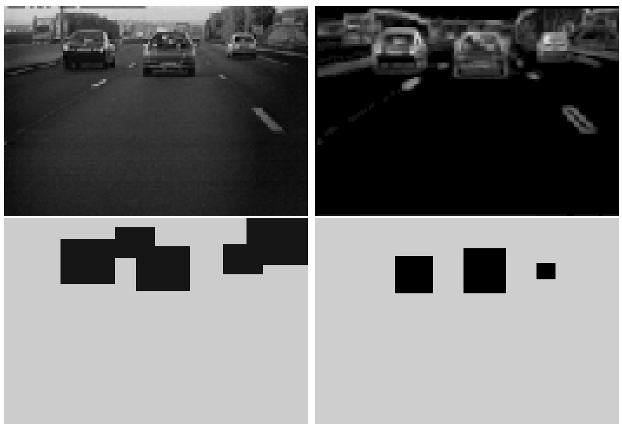

Fig. 7. Intensity, local variance , shadow detection, radar sensor

To get a stabilization over time, the segmented frame t (with additional noise) is fed back as additional feature for the segmentation task of frame $(t+1)$.

The local variance is used as a second integrating feature to increase the redundancy of the segmentation task.

Fig. 8. Sensor fusion for segmentation

The shadow detection algorithm detects the vehicle on the basis of the light conditions in natural traffic scenes [2]. The algorithm searches for horizontal orientated shadows (positive grey tone gradient) on the lane. Above these regions vehicles can be detected with a high probability [13]. With this information a picture based, binary representation with hypotheses of vehicles is calculated.

The radar system detects up to three objects in front of the car and tracks them with an electronic pivotable pencilbeam [14]. The distance is determined from the impulse run time to the radar target. The direction to the radar target is calculated by the dependence of the angle position of the well bundling antenna.

The radar data supply additional information, which is not contained in camera pictures. Furthermore the radar information is reliable also with poor view visibility (rain, fog) and leads to a more robust segmentation. In a sensor picture with the same pixel oriented base representation as the feature pictures (binary) of the base algorithms, the positions of the vehicles are represented as squares.

The fusion of the features $\mathbf{w}$ (feedback over time $w_1(x,y)$, local variance $w_2(x,y)$, shadow detection $w_3(x,y)$, and radar information $w_4(x,y)$) can be extended in this way to a multi sensor fusion. The vector $\mathbf{f_3}(x,y)$ is defined as

$$\mathbf{f_3}(x,y) = \sum_{(i,j)\in R} \mathbf{w}(i,j). \tag{7}$$

whereby R describes the local neighborhood (9×9) of a pixel (x,y). A 16-dimensional input vector

$$\mathbf{f}(x,y) = (\mathbf{f_1}(x,y)^T, \mathbf{f_3}(x,y)^T, x, y)^T \tag{8}$$

for the coupling net (MLP with a 16-5-1 structure) results for each pixel.

In figure 8 a sequence (every 250th frame) of 2750 frames with the result of the fusion process is shown. All vehicles are segmented, even if more than three vehicles (three radar beams) have to be detected. The system is robust with respect to wrong input of one of the coupled sensors or algorithms [14].

IV. Discussion

A fusion process of different sensors and algorithms for segmentation is presented. The operability on the basis of three examples is demonstrated. By the use of a broad feature spectrum (edges, texture, color, radar information) an improved segmentation in natural traffic scenes is achieved. The flexibility of the fusion process permits the simple integration of color information, as well as other procedures. In the third example a sensor fusion task with optical and radar sensors is solved. In order to eliminate occurring false segmentation a fusion with other sensor data (e.g. lidar or infrared sensors [15]) may be applied.

The selection of a neural net permits an adaptation on resuming tasks (e.g. innercity traffic) by the exchange of the net weights. An on-line learning is possible.

The presented fusion net is part of a total system. In order to achieve a reliable analysis of traffic scenes, segmented areas must be classified (background, obstacle, vehicle) and tracked over time. The MLP-structure can also be assigned to these processing tasks. The results of these tasks can be integrated easily into the fusion net for further stabilization of the segmentation result.

References

[1] W. von Seelen et al., "Image Processing of Dynamic Scenes," Internal Report IRINI 97-14, Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany, July 1997.

[2] U. Handmann, T. Kalinke, C. Tzomakas, M. Werner, and W. von Seelen, "Computer Vision for Driver Assistance Systems," in *Proceedings of SPIE Vol. 3364*, Orlando, 1998, pp. 136 – 147, SPIE, Session Enhanced and Synthetic Vision 1998.

[3] U. Handmann and T. Kalinke, "Fusion of texture and contour based methods for object recognition," in *ITSC'97, IEEE Conference on Intelligent Transportation Systems 1997*, Boston, 1997, IEEE, Session 35: Intelligent Vehicles: Vision(3).

[4] R. R. Brooks and S. S. Ivengar, *Multi-Sensor Fusion*, Prentice Hall PTR, Upper Saddle River, New Jersey, 1998.

[5] C. Goerick, "Local Orientation Coding and Adaptive Thresholding for Real Time Early Vision," Internal Report IRINI 94-05, Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany, June 1994.

[6] C. Goerick, D. Noll, and M. Werner, "Artificial Neural Networks in Real Time Car Detection and Tracking Applications," *Pattern Recognition Letters*, 1995.

[7] T. Kalinke and W. von Seelen, "Entropie als Maß des lokalen Informationsgehalts in Bildern zur Realisierung einer Aufmerksamkeitssteuerung," in *Mustererkennung 1996*, 1996, pp. 627–634.

[8] S. Haykin, *Neural Networks*, Macmillan College Publishing Company, New York, 1994.

[9] B. V. Dasarathy, *Decision Fusion*, IEEE Computer Society Press, Los Alamitos, 1994.

[10] G. Lorenz, "Erzeugung eines adaptiven Objektformmodells unter Ausnutzung des Farbverteilungskontrastes," in *Proceedings 9. Aachener Kolloquium 'Signaltheorie' Bild- und Sprachsignale*, RWTH Aachen, 1997, pp. 179–182.

[11] W. Ritter, F. Stein, and R. Janssen, "Traffic Sign Recognition Using Colour Information," *Mathl. Comput. Modelling*, vol. 22, no. 4–7, pp. 149–161, 1995.

[12] J.D. Foley, A. van Dam, S.K. Feiner, and J.F. Hughes, *Computer Graphics: Principles and Practice*, Addison-Wesley Publishing Company, Inc., 1990.

[13] F. Thomanek, E.D. Dickmanns, and D. Dickmanns, "Multiple Object Recognition and Scene Interpretation for Autonomous Road Vehicle Guidance," in *Proceedings of the Intelligent Vehicles '94 Symposium, Paris, France*, 1994, pp. 231–236.

[14] T. Schnitger and U. Handmann, "Fusion von Bildanalyseverfahren mittels einer neuronalen Kopplungsstruktur," Internal Report IRINI 98-01, Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany, Apr. 1998.

[15] L. C. Ramac, M. K. Uner, P. K. Varshney, M. Alford, and D. Ferris, "Morphological filters and wavelet based image fusion for concealed weapons detection," in *Proceedings of SPIE Vol. 3376*, Orlando, USA, 1998, SPIE, Session Sensor Fusion 1998.