

Pedestrian Detection with Radar and Computer Vision

**Milch, S., Behrens, M., smart microwave sensors GmbH,
Braunschweig**

Abstract

This paper presents a method for detecting pedestrians on-board a moving vehicle. The perception of the environment is performed through the fusion of an automotive radar sensor and a monocular vision system. The fusion uses a two-step approach for efficient object detection. In the first step, a target-list is generated from the radar sensor. The items in the list are hypotheses for the presents of pedestrians. In a second step, hypotheses are proved by the vision system. This method achieves very large speed-ups compared to a sole image processing solution.

Introduction

Pedestrians are one of the most valuable traffic participants. In Germany, more than 39.000 pedestrians were injured in 1999 alone, due to collisions with vehicles. Of these, more than 900 were deadly injuries [7].

The aim is to develop assistant and safety systems to avoid these accidents or at least minimize their severity.

To detect pedestrians with an artificial system is difficult for a number of reasons. The main challenge for a vision-based pedestrian detector is the high degree of variability with the human appearance due to articulated motion, body size, partial occlusion, inconsistent cloth texture, highly cluttered backgrounds and changing lighting conditions. Moreover, the applications, to protect pedestrians, defines hard real-time requirements and rigid performance criteria.

System outline

The system consists of an automotive radar and a video sensor. Both sensors have different properties (Table 1).

The idea of combining multiple inputs to infer information about the environment is very natural. It is done by humans in everyday live, since we are all the time combining acoustic, visual and tactile information to get more reliable knowledge about the world around us. Sometimes it is not possible to derive the information needed for a particular task from one single sensor. Furthermore, there is no perfect sensor, so it is reasonable to make use of the favorable properties of one sensor and to suppress the disadvantages by applying a smart combination scheme. Sensor

fusion means a very wide domain and it is difficult to provide a precise definition. We use the following definition based upon the work of Wald [6]:

“... data fusion is a formal framework in which are expressed means and tools for the alliance of data originating from different sources. It aims at obtaining information of greater quality; the exact definition of «*greater quality*» will depend upon the application.”

	Radar	Video
Result of measurement	List of reflection-points (range, velocity, angle, RCS ¹)	Grayscale matrix (brightness distribution)
Sensor principle	Active	Passive
Data rate	Low	High
Object detection	Clustering of reflection-points (without model)	Knowledge based interpretation (with model)
Object properties	Location, velocity, RCS	Model depending

Table 1: Sensor Properties: Radar and Video

Here, quality has not a very specific meaning. It is a generic word denoting that the resulting information is more satisfactory for the “customer” when performing the fusion process than without it. In our case, the aim is increased confidence. In addition, we must consider performance issues related to computational complexity and accuracy. The application require computation to be performed on-line under real-time. A fusion process can take place on different hierarchical levels. 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, in order to use the advantage of data reduction compared with a fusion on data level.

Figure 1 outlines the intersection area from both sensors. Each sensor measures two dimensions of a three dimensional world.

Different architectures are possible for sensor fusion. Mostly common is a parallel combination of sensors to extract feature vectors for the observed objects. In this application a sequential analysis of the feature vector speeds-up processing. Sequential analysis leads to a hierarchical fusion architecture. The radar sensor generates a list of objects. For each object distance, angle and radar cross section (RCS) are extracted. The radar sensor detects objects without an explicit object-

¹ RCS – Radar Cross Section: A measure of the reflective strength of a radar target. Usually represented by the symbol σ , measure in square meters, and defined as 4π times the ratio of the power per unit solid angle scattered in a specified direction to the power per unit area in a plane wave incident on the scatterer from a specified direction. The value depends on shape, size, material properties and aspect angle (wave – object).

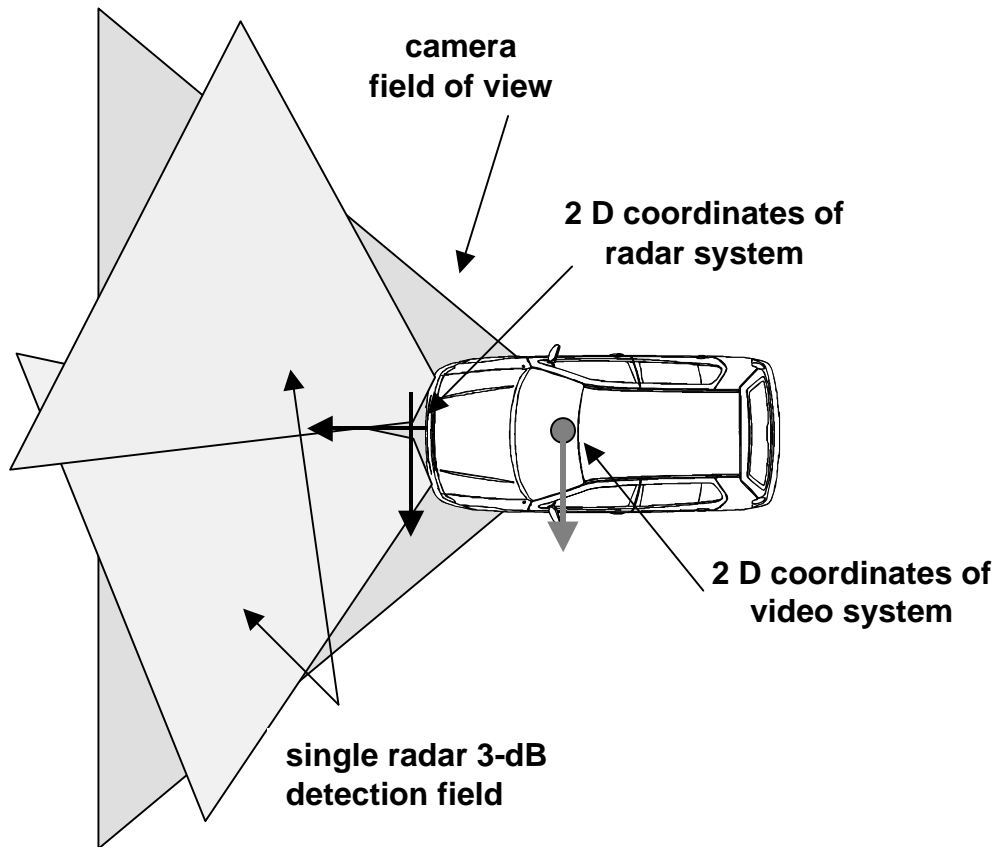


Figure 1: Field of view

model. In addition, computational complexity is rather low. In a pre-selection phase, the pedestrian candidates are filtered: only the ones that satisfy specific constraints on the speed, RCS and size are selected as hypotheses. These hypotheses are the input for the signal-processing module of the vision sensor (Figure 2).

Unlike radar, computer vision needs object-models for the objects of interest. These models contain in most cases parameters to cover changes of appearance [4]. Object detection is a search of maximum of a similarity function. The dimension of search-space is in worst case equal to the dimension of the parameter vector of the used model. With hierarchical sensor combination dimension of search-space is reduced drastically.

Radar System Overview

The radar system consists of two individual radar sensors with slightly overlapping field of view and a central processing unit that fuses the radar sensor data and vehicle information like velocity and steering. Further, it tracks the data over time and selects object data for the hypotheses list.

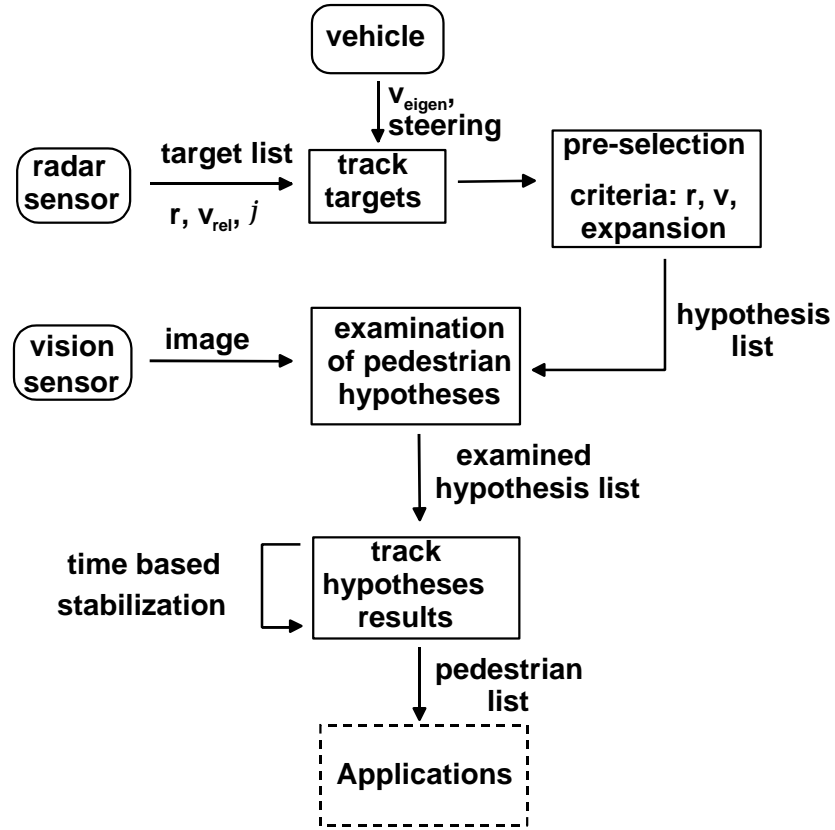


Figure 2: Topology of the pedestrian detection system

The advantage of using a radar system is that the functionality is nearly uninfluenced by weather, day/night conditions and pollution. The installation could be done invisible behind a bumper, so that the vehicle contour and design is not influenced. Radar measures runtime, power and Doppler frequency shift of electromagnetic waves, transmitted from the sensor and reflected back from objects in the field of view. Runtime is a measure for the distance and Doppler frequency shift for the velocity of an object. The angle between the radar sensor bearing and the object is measured comparing phase values of the received signal. An overview on the sensor characteristics is given in Table 2.

The radar sensor measures only point targets or sub reflectors of an object with great expansion. Only objects like cars, trees, traffic signs, bicycles and human beings are detected. Street and pavement are not seen due to the lack of significant reflectors.

In [5] the so-called radar equation is given

$$P_{receive} = P_{transmit} \cdot \frac{G_1(\mathbf{j}) \cdot G_2(\mathbf{j})}{(4\pi)^3 \cdot r^4 \cdot I^2} \cdot \mathbf{s}, \quad (1)$$

where r is the measured range and G_1, G_2 the antenna gain depending on the object angle \mathbf{j} . With the received power $P_{receive}$ and (1) an estimation of RCS \mathbf{s} for each detected object could be calculated. RCS is used to differentiate between objects, e.g. pedestrians have a typical RCS between 0.01 m² and 1 m² at 24 GHz while cars

have values between 0.1 m² and 1000 m². Another criterion is object velocity. Pedestrians have a maximum velocity less than 10 m/s (running), normally less than two m/s, while cars and bikes could have much higher velocities. Classification between point targets (e.g. traffic sign, pedestrian) and area targets like cars is done by clustering object data using correlated range, velocity and angle. The hypotheses list is given by the intersection set of objects that are point targets and fulfils the RCS and velocity criteria for pedestrians.

Carrier Frequency	24.125 GHz ($\lambda = 12,4$ mm)
Transmit Power	10 mW
Size (WxHxD)	75 mm x 90 mm x 35 mm
Field of View	Azimuth: 50° ($\pm 25^\circ$) Elevation: 16° ($\pm 8^\circ$)
Maximum Range	up to 40 m
Used Measurement Range	$r = [0.1 \dots 20]$ $v = [-80 \text{ m/s} \dots 80 \text{ m/s}]$ $\mathbf{j} = [-25^\circ \dots +25^\circ]$
Measurement Accuracy	$\tilde{r} < 0.1$ m $\tilde{v} < 0.1$ m/s $\tilde{\mathbf{j}} < 1^\circ$
Communication Interface	CAN-Bus

Table 2: Radar Sensors Properties

Hypothesis Verification by Vision Data

In this stage, the pedestrian-hypotheses from the radar system are checked with additional information from the video sensor. A hypothesis consists of a position, a velocity and a moving direction. The position is determined by range and angle, the altitude is not known. To transform points from the coordinate system of the radar sensor in the coordinate system of the vision system, the orientation of both sensors must be known in 3D world-coordinate system. A 2D point is transformed with an estimated altitude in the 3D world-coordinate system. With a camera model the projection from 3D world- to 2D imager coordinates is calculated (Figure 3).

The vision system has been designed to work with only grayscale images, either visible or near infrared. While most previous work on detecting people relies heavily on color cues, this system is designed for outdoor scenarios, and particularly for nighttime or other low light level situations. In such cases, color will not be available.

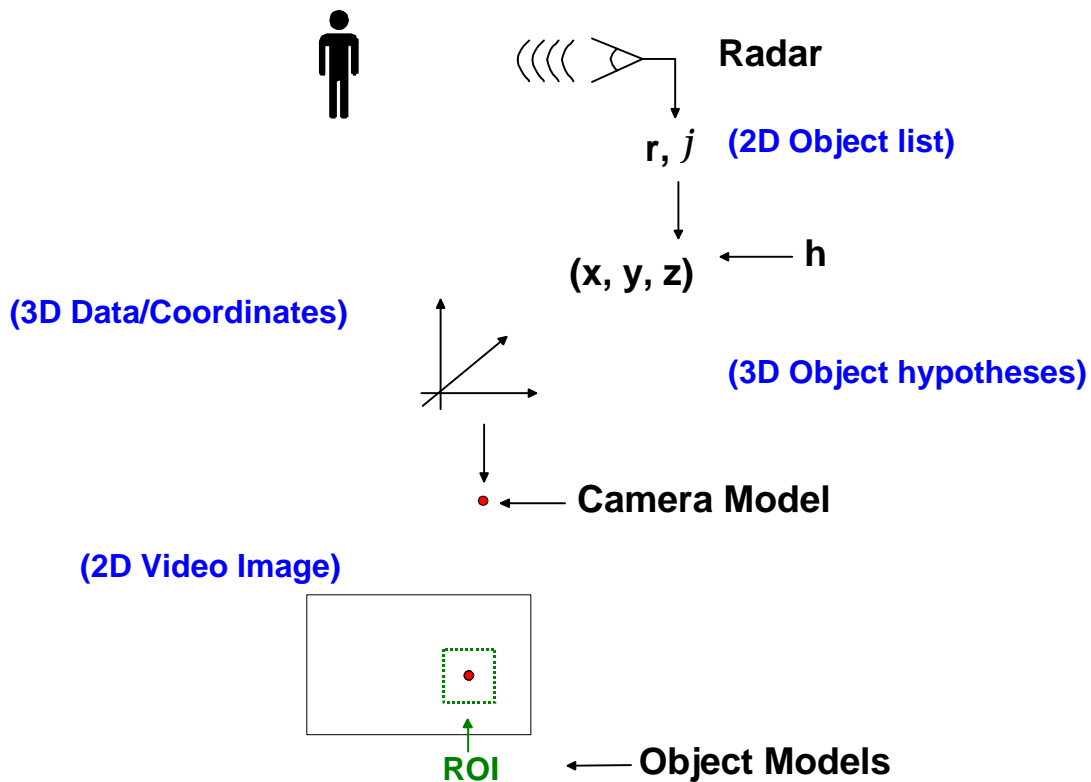


Figure 3: Transformations between different coordinate systems

The hypothesis is checked with a pedestrian model. It is essential to ask what form of representation are suitable for mediating effective computation for the perception of dynamic visual objects as pedestrians in traffic scenarios. In other words, what information is most relevant to the perception of human bodies and what form of representation enables such information to be extracted from video images and utilized under the demands of temporal and computational constraints?

A comprehensive literature review for models to represent a priori information about human appearance is given by Gavrilu [2].

We use a flexible 2D prior model of silhouette shape to recognize and track pedestrians in image sequences. Classification is performed based on shape information. Two-dimensional deformable models, also known as “active contours” or “snakes”, were originally proposed by Kass et al [3]. Snake was presented as energy-minimizing parametric closed curve guided by external forces. Because snakes do not incorporate prior knowledge about expected shapes, this approach is easily confused by other structures present in the image and occlusion.

A prior shape model incorporates useful constraints on the apparent shape of a pedestrian silhouette that allows the system to cope with missing information due to image noise, background clutter and partial occlusions. A deformable model is required to model the apparent change in shape due to pose (i.e. position of limbs etc) and viewpoint relative to the camera.

We trained a flexible shape model using pedestrian shapes extracted manually from video sequences. During training, the distribution of shape parameters was established. Two different models were trained, one for a frontal view and one for a side view of the human body.

The initial placement is performed by the hypothesis from the radar sensor. At this stage, a person is considered to be average world height and to be vertical in the image. An image fitting process is used to provide a fitness measure for the current hypothesis. The fitness measure is the percentage of the contour that is locked onto a significant image feature. If the fitness for a hypothesis is above a threshold then the hypothesis is accepted. Otherwise, the hypothesis is rejected and no longer examined.

Conclusions and Outlook

In [1] static methods for pedestrian protection are examined. A sensor system, that is able to detect pedestrians, makes it possible to develop an active protection system. Active systems have advantages for cars with tight engine packaging, in this case an "active hood" is able to decrease impact severity. The feasibility of the suggested system was shown, performance analyzes, enhancements and optimization are still under examination.

References

- [1] Brown, G., "Headlight Design Changes Resulting from Proposed Pedestrian Protection Requirements", In: PAL Progress in Automobile Lighting Vol. 5, Symposium Proceedings, Darmstadt University of Technology, 1999
- [2] Gavrilu, D., "The visual analysis of human movement: A survey", Computer Vision Image Understanding, Vol. 73, No.1, pp 82-98, 1999
- [3] Kass, M., Witkin, A., Terzopoulos, D., "Snakes, Active contour models", First International Conference on Computer Vision, pp. 259-268, IEEE, Computer Society Press, 1987
- [4] Milch, S., " Videobasierte Fahreridentifikation in Kraftfahrzeugen ", Dissertation Universität Darmstadt: Fachgebiet Lichttechnik, Utz-Verlag, München, 2001
- [5] Skolnik, M. I., „Introduction to Radar Systems“, Chapter One, McGraw Hill, 1981
- [6] Wald, L., "A European proposal for terms of reference in data fusion", In: International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 7, pp 651-654, 1998
- [7] BASt: "Straßenverkehrsunfälle in Deutschland", <http://www.bast.de/>, 2001