The FRESNEL Program: Fusion of Radar and Electro-optical Signals for Surveillance on Land

Arne Theil, Leon J.H.M. Kester, Sebastiaan P. van den Broek, Philip van Dorp, Ronald van Sweeden

TNO Physics and Electronics Laboratory
P.O. Box 96864
2509 JG The Hague
Netherlands

ABSTRACT

Area surveillance for guarding and intruder detection with a combined camera radar sensor is considered. This specific sensor combination is attractive since complementary information is provided by the respective elements. Thus, a more complete description of objects of interest can be obtained. Several strategies to fuse the data are discussed. Results obtained with 'live' experiments are presented. When compared to camera only, a significant reduction of the number of false tracks is achieved.

Keywords: sensor fusion, tracking, surveillance, surface picture

1. INTRODUCTION

In this paper, area surveillance with a combined camera radar sensor ('cameradar') is considered. This topic is being investigated at TNO Physics and Electronics Laboratory (TNO-FEL) within a program referred to as *FRESNEL*: Fusion of Radar and Electro-optical Signals for Surveillance on Land. The application is automated guarding and intruder detection.
The particular sensor combination is attractive since the individual components provide information that is partly complementary, one speaks of 'dissimilar sensors.' Generally, a camera provides highly accurate angular information (azimuth, elevation), while a radar excels in range and —provided the system is coherent— range-rate (radial velocity). Thus, a more complete description of objects of interest ('targets') can be achieved by combining the data. As a consequence, an enhanced classification capability and a reduced number of false alarms per unit of time can be noted among the benefits. Such advantages are highly desirable when a surveillance task needs to be automated.
The sequel of the paper is organised as follows. In section 2 fundamental aspects of sensor fusion when dealing with dissimilar sensors are outlined. Subsequently, several architectures that are applicable to fuse data from dissimilar sensors are discussed (section 3). The experimental setup is described in section 4, results are presented in section 5. Finally, conclusions are formulated and references are given.

2. FUSION OF DATA FROM DISSIMILAR SENSORS

The approach one should follow to fuse data from dissimilar sensors is not trivial. As a consequence of the very dissimilarity of the individual sensors, low level fusion applied to non-processed ('raw') input data, e.g. image fusion, is not applicable since the data don't match. To illustrate this statement, two examples of scenes observed by a cameradar are provided in figure 1 (next page). Notice that both sensors are equipped with algorithms to extract features of possible targets. An output message containing values of features is referred to as a contact. For the specific scenes and sensors as depicted in figure 1, the quality of the extracted features is summarised in a qualitative manner in table 1.

Signal Processing, Sensor Fusion, and Target Recognition X, Ivan Kadar, Editor,
Proceedings of SPIE Vol. 4380 (2001) © 2001 SPIE · 0277-786X/01/$15.00

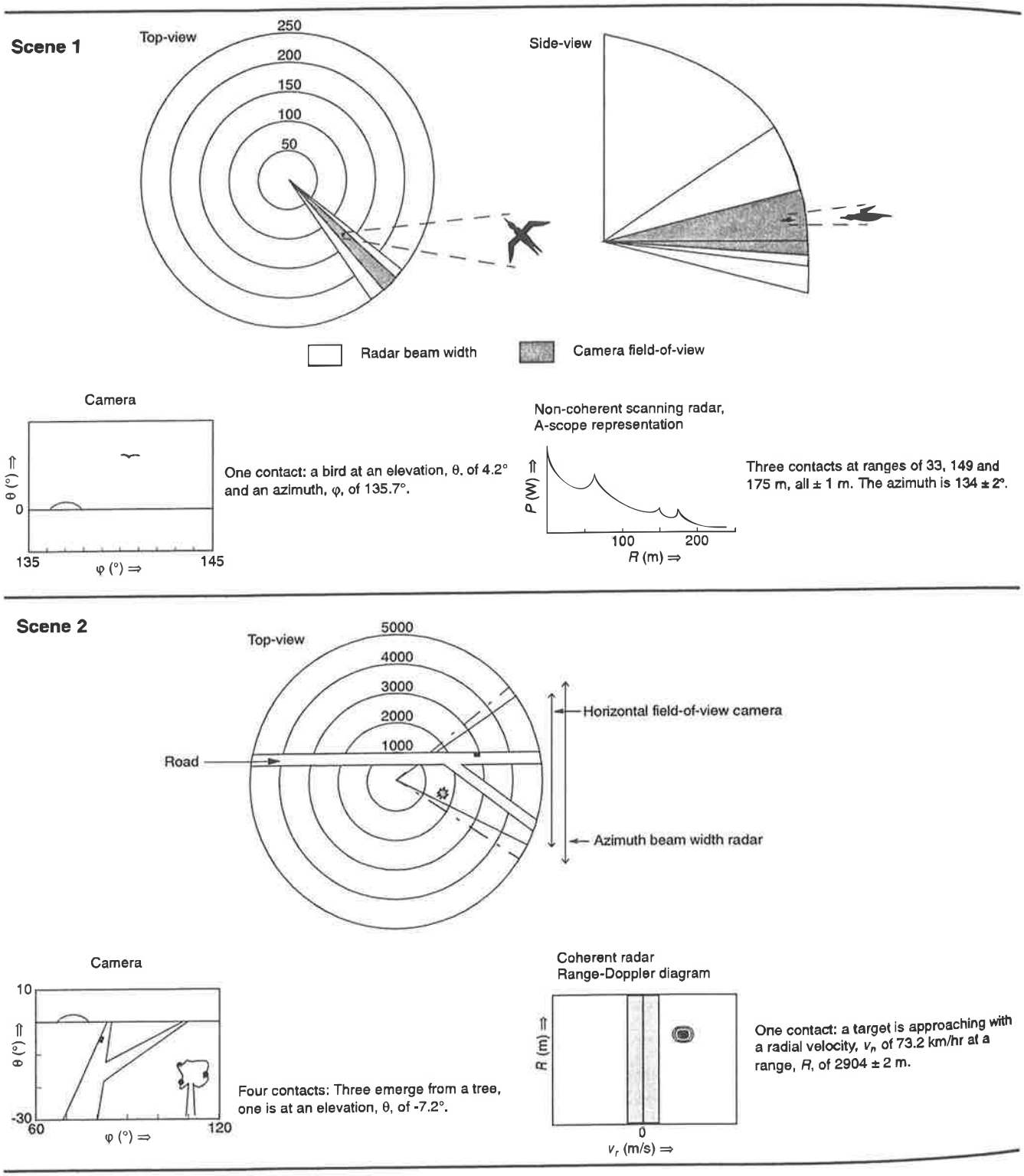

Figure 1: Examples of camera and radar data to illustrate observations from dissimilar sensors.

Table 1: Overview of extracted features and their quality for the scenes depicted in figure 1. '+' = good, '-' = poor.

	Feature	Camera	Radar
Scene 1	Range		+
	Azimuth	+	-
	Elevation	+	-
	Class	+	
Scene 2	Range		+
	Azimuth	+	-
	Elevation	+	-
	Range-rate		+

The table shows that additional information is needed to associate the contacts produced by the two sensors with each other. For scene 1, knowledge on the maximum bird size favours the option that the bird is at a distance of 63 m. (The bird covers 1° in azimuth. For an Albatross with 2.5 m wingspan this relates to a distance of 143 m.) For scene 2, association can be achieved by assuming that the targets are bound to the earth's surface. Range information can then be derived from the elevation that is provided by the camera, so that a link to the radar information is established. In case none of the radar contacts matches a specific camera contact in range, it must be concluded that either the camera contact represents a false alarm or an airborne target, or that the radar has failed to observe the target, i.e., a so called 'missed detection.'
These examples illustrate that fusion of data from dissimilar sensors relies on a mechanism to associate the data. If such a mechanism can't be found, fusion is not possible. Furthermore, contact extraction per sensor is required in order to obtain one or multiple features that provide the key for fusion.

3. SENSOR FUSION ARCHITECTURES

3.1 Contact fusion, fuse-while-track

In case sensors do not operate synchronously, i.e., they produce contacts at different points in time, the so-called fuse while track architecture can be applied, see figure 2.

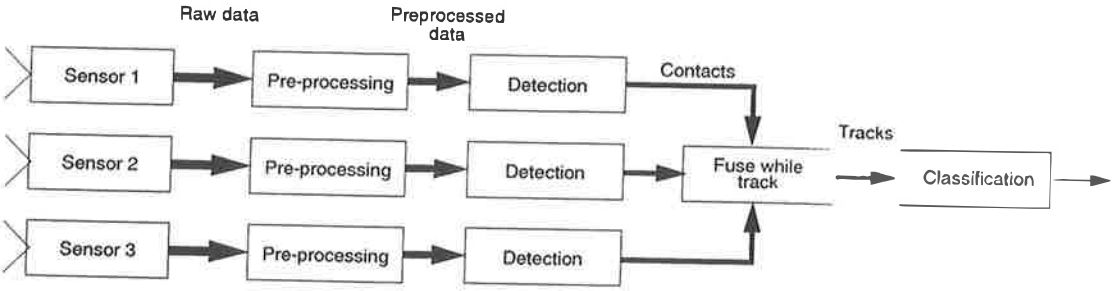

Figure 2: Fuse-while-track architecture. The width of the arrows corresponds with the data rate, expressed in bits/s.

A central tracking process receives contacts from several sensors and associates these with tracks: contact-to-track-association. In case of dissimilar sensors, the tracker must be able to handle contacts with deviant content.
The processes tracking and classification are in fact optional. Tracking is generally highly effective to reduce the number of false alarms (contacts). Furthermore, the quantities that are estimated by the track filter (e.g. kinematic properties) can be exploited to establish or refine the classification. Hence, we prefer to place the classification process at the end of the signal processing chain.

3.2 Track fusion

A further decentralisation of processing is intrinsic to the track fusion architecture, as illustrated in figure 3. The fusion process fuses tracks that are produced by the individual sensors: sensor-track-to-track-association. Owing to the false alarm reduction achieved by the individual trackers, the data rate on the input side of the fusion process is relatively low compared

to the architectures previously discussed. Several authors have argued the inferiority of track fusion compared to contact fusion, due to mutual correlation of tracks.

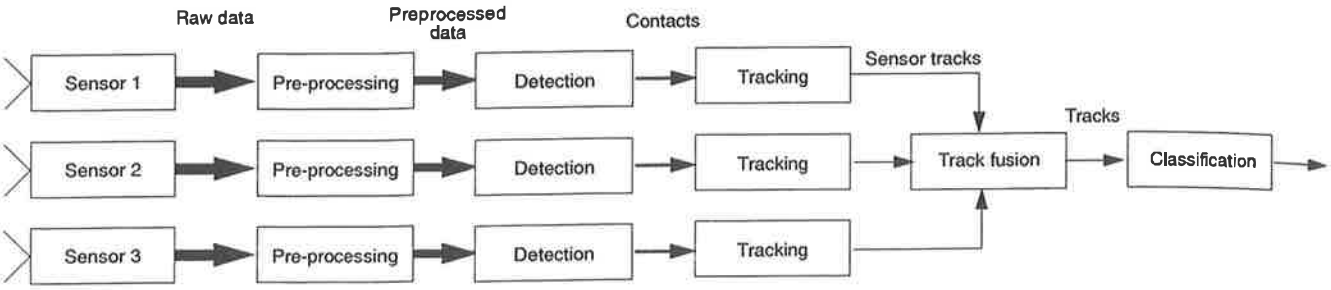

Figure 3: Track fusion architecture.

3.3 Hybrid architectures

Two examples of hybrid architectures are given in the figures 4 and 5. In the upper scheme, local tracking is applied in one of the channels merely to establish a reduction in the number of contacts; only contacts that associate with local tracks are being offered to the fuse-while-track algorithm. This recipe is attractive in case certain sensors produces significantly more false contacts than the other sensors. A mixture of fuse-while-track and track fusion is shown in figure 5.

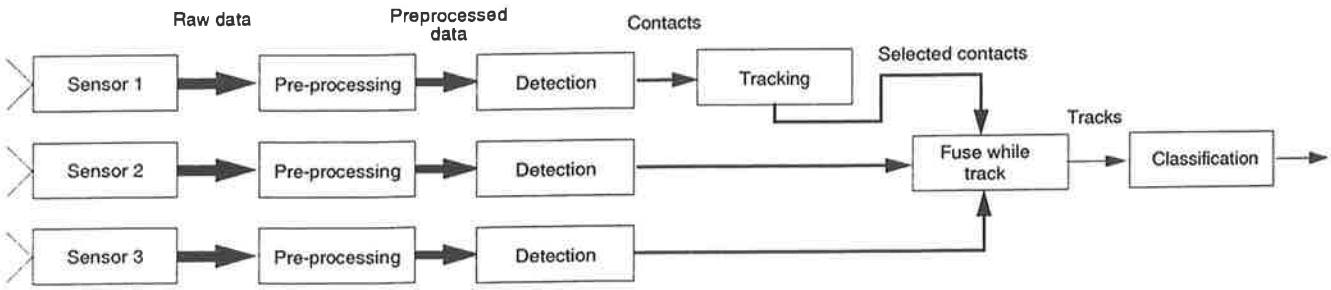

Figure 4: Hybrid sensor fusion architecture.

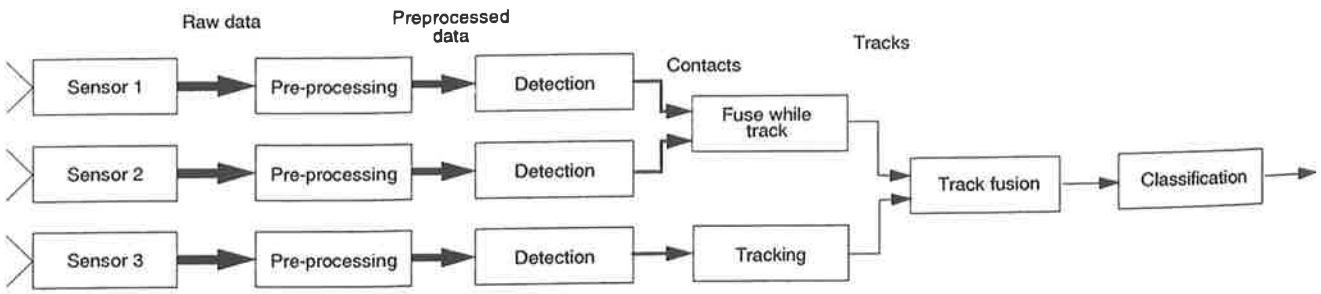

Figure 5: Hybrid sensor fusion architecture.

4. EXPERIMENTAL SETUP

In this communication results are presented obtained with a cameradar that comprises a visual light b&w CCD camera and a coherent X-band FM-CW radar, closely co-located, see figure 6. The cameradar operated in a non-scanning mode.

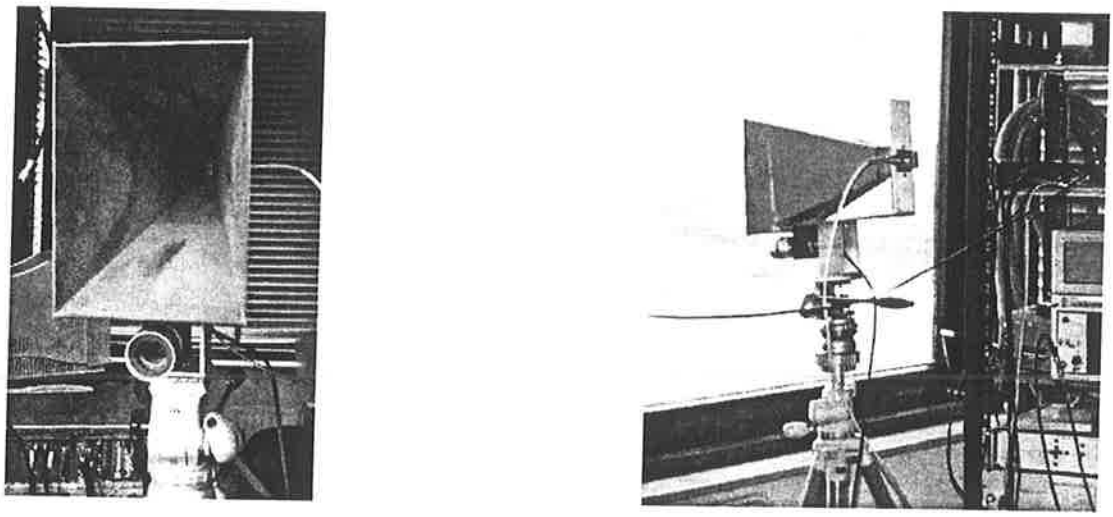

Figure 6: Cameradar used for the FRESNEL experiments. An X-band horn antenna is mounted above a visual light CCD camera providing b&w images.

The radar was equipped with a single rectangular horn antenna, offering beam widths of 12° in azimuth and 8.5° in elevation at horizontal polarisation. The camera's field-of-view was 5.5° × 4.2° (azimuth × elevation). It produced 8-bit b&w images, sized 748 × 572 pixels, at a rate of 25 Hz. All raw data was recorded on hard disks, processing was done off-line.
The cameradar was positioned on a tower at an altitude of 25 m. Recordings were made of people and cyclists, typically some 200 to 500 m from the foot of the tower. Both single and multi target data was gathered. An impression of raw camera data and pre-processed radar data in case of a single target scenario is given in figure 7.

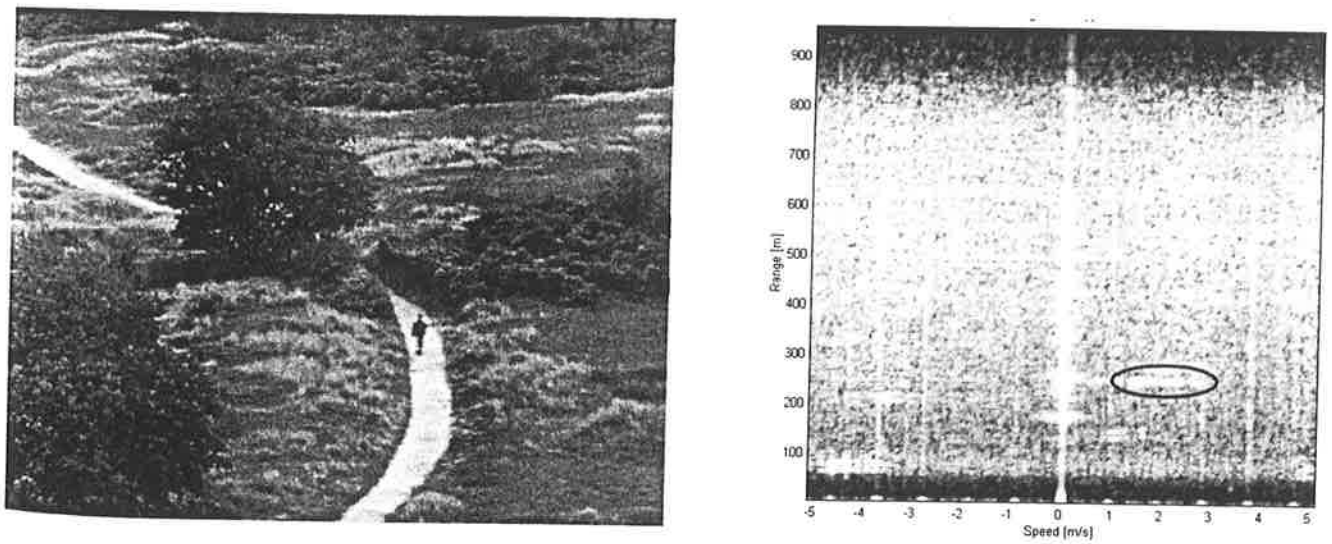

Figure 7: Data provided by the cameradar. Left: visual light b&w image. Right: range-Doppler diagram; responses from a walking person are encircled.

5. RESULTS

5.1 Introduction

In this section results obtained from a rather straightforward scenario in which an inbound walking person is observed during approximately 30 s are presented. The objective of the analysis is as yet predominantly to assess the possible surplus value of sensor fusion. To quantify this, we mainly consider track statistics, particularly the number of false tracks, since we aim at automated surveillance. For an overview of measures of performance to characterise sensor fusion effectiveness, we refer to Theil and Kester [1].

5.2 Camera contact extraction

The detector used for the extraction of contacts from the visual light images, detects movement by comparing the grayvalue of each pixel with a threshold, T, that is computed from the previous 25 frames, according to:

$$T = \mu + k\sigma \tag{1}$$

with μ and σ the mean grayvalue and its standard deviation and k a given constant. Pixels in which a detection occurs are then clustered. Clusters with less than a given number of pixels are removed. Contacts are extracted from the remaining clusters. With this method of moving background estimation, an object moving into the field of view of a pixel will cause its average and standard deviation to be changed, reducing the contrast and increasing the threshold. As a result, only the edge of an object will be detected. This effect can be avoided by disregarding pixels labelled as detections in the background estimation. Because of this, standard deviation estimates are slightly lower, but it allows the detection of the whole object, even when it stops moving. This processing appeared to be well capable in detecting moving objects every frame, provided they are not occluded.

In figure 8 the camera contacts during the single target experiment are presented. The left diagram shows the contact positions in the pixel co-ordinate system of the camera-plane. To ease a comparison with radar data, the right diagram shows ground range versus time.

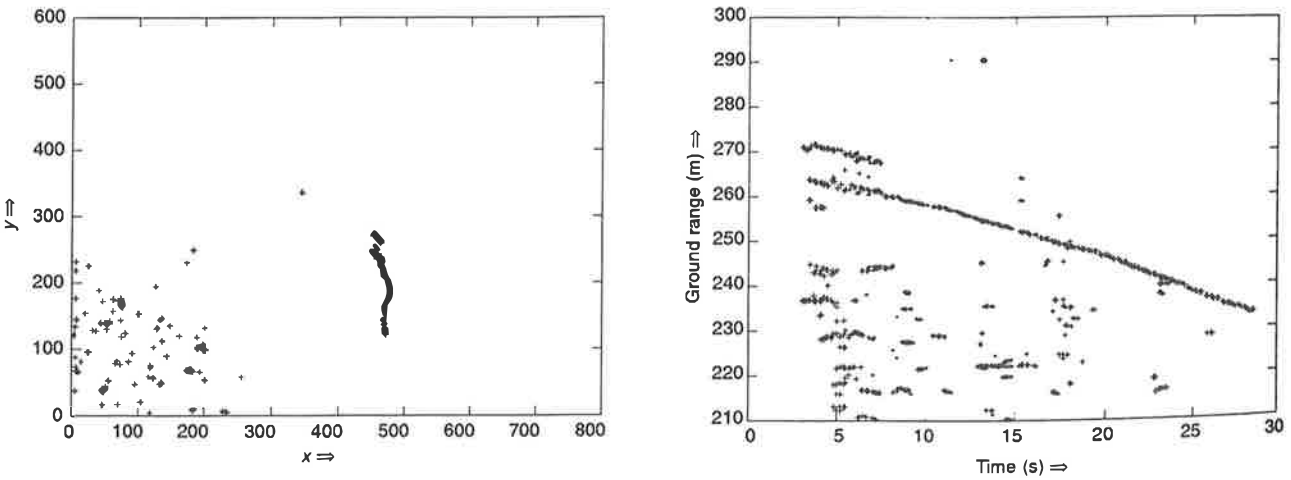

Figure 8: Camera contacts obtained from the single target experiment. 208 frames during 25.5 s delivered a total of 563 contacts. Left diagram: contact positions depicted in the pixel co-ordinate system. Right diagram: ground range versus time.

Responses from the walking person can easily be identified in both diagrams. One may notice that in approximately 20% of the frames, two or three contacts are produced from the person. Non-target contacts are attributed to moving parts of overgrowth, such as trees and shrubs. Rather than the full frame-rate, a rate of 8.2 frames per s was employed. This resulted in an average of 2.7 contacts per frame.

5.3 Radar contact extraction

The pre-processing entailed a 1024-point range-FFT (real-valued input data) and 128-point Doppler-FFT's for 512 range cells (complex-valued input data). This resulted in range-Doppler cells of size 5 m × 0.08 m/s (depth × range-rate). Subsequently,

a detection algorithm based upon cell averaging - greatest of (CAGO) CFAR is applied to the data in each Doppler-bin. Next, clustering in range and in Doppler is carried out to identify neighbouring cells.
In figure 9 the radar contacts during the single target experiment are presented.

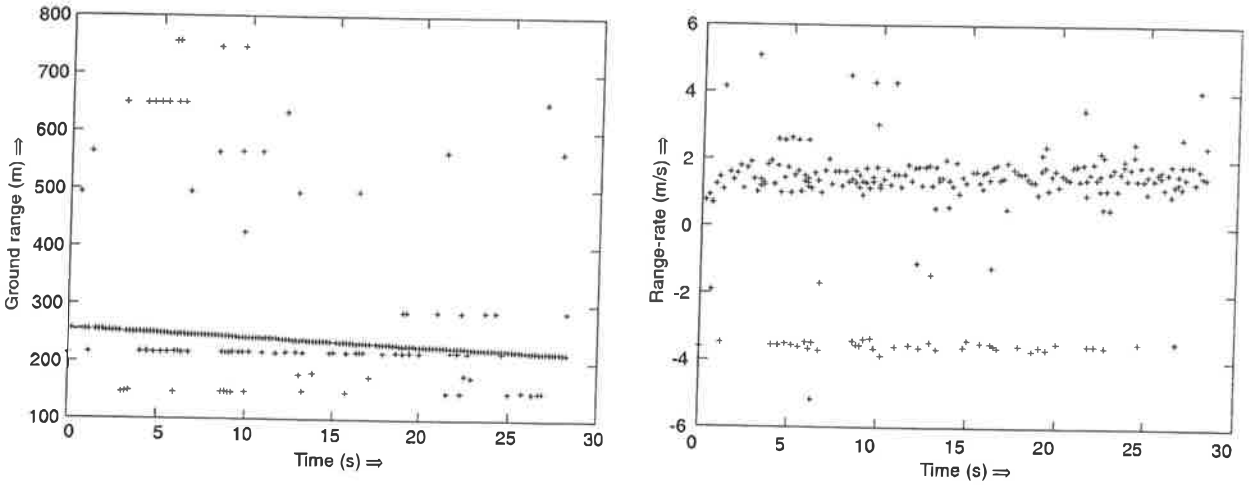

Figure 9: Radar contacts obtained from the single target experiment. 150 measurements during 28.4 s delivered a total of 236 contacts. Left diagram: ground range versus time. Right diagram: range-rate versus time.

Notice that the radar is capable in detecting the person at nearly every measurement. Besides, the phenomenon of multiple target contacts per measurement does not occur. The estimated range-rate of the person, shown in the right diagram at approximately 1.5 m/s, is somewhat irregular. This is caused by the movement of limbs. An average of 1.6 contacts per measurement is produced. Non-target contacts are mainly caused by artefacts of the radar. We estimate that only 10% of these contacts originate from moving parts of growth.

5.4 No fusion

In figure 10 information of tracks that are obtained on camera contacts only is shown. A Kalman track filter has been applied that assumes target motion according to a constant velocity kinematic model, see for instance Blackman and Popoli [2]. Contact-to-track association is performed on basis of nearest neighbour. The diagram at the left shows tracks projected on the horizontal plane at the foot of the tower. Ground range versus time is shown on the right.

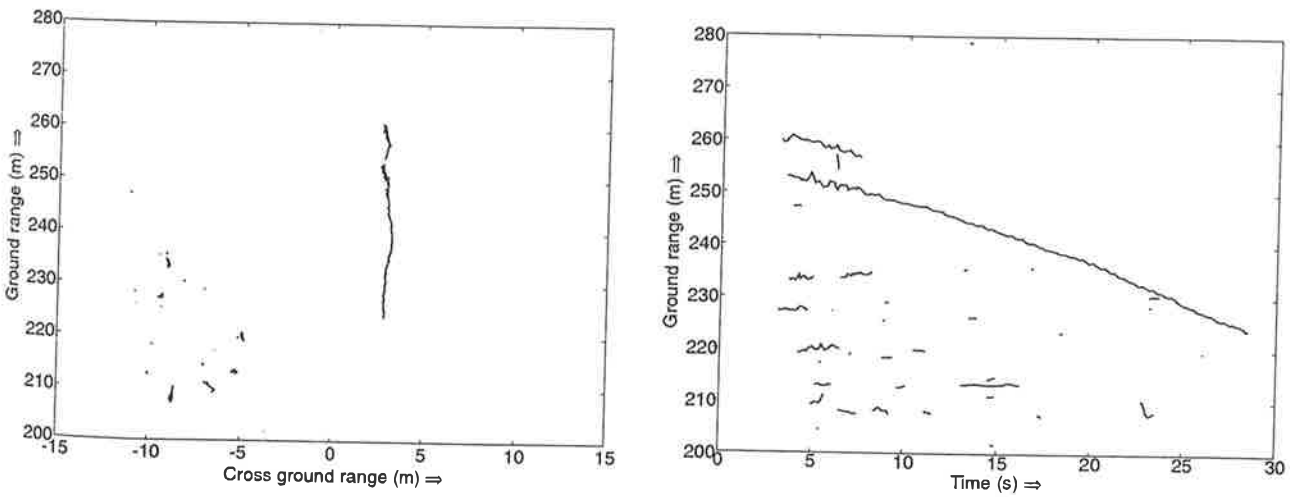

Figure 10: Tracks in the single target scenario, camera data only.

Target tracks appears vertically, in the middle of the left diagram. False tracks, emerging from the crown of the tree in the foreground of the observed scene (see figure 7), appear at the lower left. The fact that the target gives rise to three tracks is more easily to be seen in the right diagram. Obviously, a clean picture, showing only target tracks, can be obtained by applying a threshold on the lifetime of tracks. A total of 27 tracks are generated, the false track rate is 0.94 s^{-1}.
The result of range-only tracking applied to the radar data is shown in figure 11. This tracking algorithm does not take into account the measured range-rate. A 2-out-of-2 track initiation criterion has been applied. Not surprisingly, a single target track is generated. A total of 7 tracks are generated, the false track rate is 0.21 s^{-1}.

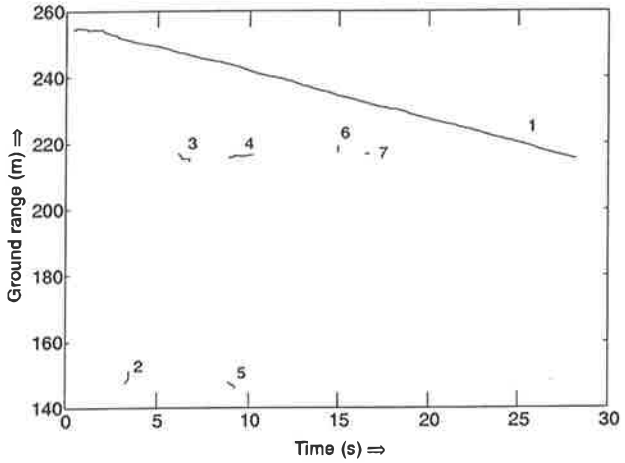

Figure 11: Tracks in the single target scenario, radar data only.

5.5 Contact fusion, fuse-while-track

Results obtained with fuse while track are depicted in figure 12. Tracks are initiated on basis of radar contacts. For details of the actual fusion algorithm we refer to Kester and Theil [3] (this conference). The irregularities shown in the left diagram are caused by the lack of azimuth information (provided by the camera) in the initial stage of a track's lifetime, zero cross range is then assumed.

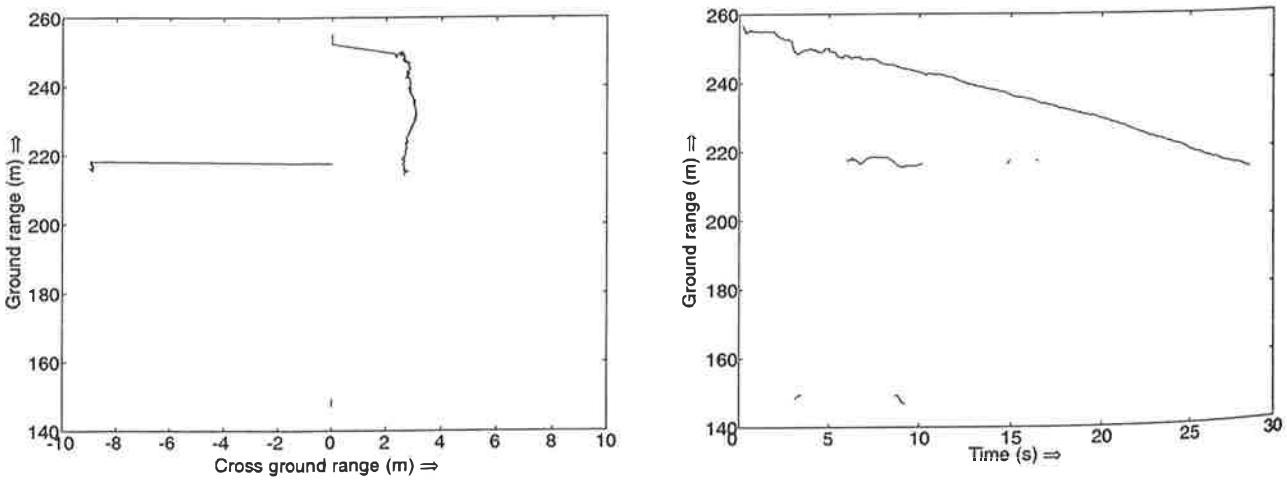

Figure 12: Tracks in the single target scenario obtained with fuse-while-track.

Compared to the result obtained with radar contacts only, one track fewer is generated, the false track rate now being 0.18 s^{-1}. In fact, the fusion remedies a track break that occurred in the radar only case (in figure 11: the tracks labelled 3 and 4).

6. CONCLUSIONS

In this paper, beneficial effects of fusion of contact data from a camera and a coherent radar has been demonstrated. Compared to the camera-only system, the number of false tracks is significantly reduced, due to combining with radar. Furthermore, connecting extracted target features from both sensors holds out the prospect of enhanced classification capabilities. It is expected that additional improvements in the fusion process can be achieved by exploiting the range-rate information provided by the radar.

Obviously, the scenario that provided data for this communication can be characterised as rather straightforward. Consequently, tracking results obtained with the individual sensors can be qualified as satisfactory as well. It is, however, expected that the benefit of sensor fusion will live up more to its promise under unfavourable circumstances. For instance, adverse weather conditions and occlusion by growth will particularly affect the performance of the camera.

A further analysis will be undertaken in the near future to obtain a full assessment of the capabilities of a combined camera radar sensor for guarding purposes.

REFERENCES

[1] A. Theil, L.J.H.M. Kester, *On Measures of Performance to Assess Sensor Fusion Effectiveness*, Proceedings of the 3rd International Conference on Information Fusion, Paris, France, July 10-13, 2000

[2] S.S. Blackman, R. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House, Norwood, MA, 1999

[3] L.J.H.M. Kester, A. Theil, *Fusion of Radar and EO-sensors for Surveillance*, SPIE Conference 4380, April 2001