

# Improving object recognition using range information

Klamer Schutte

TNO Physics and Electronics Laboratory, P.O. Box 96864, 2509 JG The Hague, Netherlands<sup>1</sup>

## ABSTRACT

This paper reports on the fusion of IR imagery with range data such as obtained from a laser range finder. Both an air/sea, and land based scenario has been studied. The range information is used to calculate a priori scale information for the detection process in the IR images. The use of this scale information leads to substantial improvement of the recognition performance.

Keywords: IR images, laser range data, sensor fusion, automatic target recognition

## 1. INTRODUCTION

Many platforms within the Dutch armed forces are now, or will be in the near future, equipped with both imaging and range indicating sensors. The imaging sensors can be IR or visual cameras. The range indicating sensors are laser range finders or radar. The spatial and / or temporal resolution of the imaging sensors is high compared to the range indicating sensors.

The research described in this report investigates into the fusion of such imaging and range sensors. The central idea is how to combine the high resolution imagery with the range data. As such, the preferred sensor within this research is the one delivering the imagery. The range indicating sensor is additional. This is not the case in all operational systems. When the imaging sensors are combined with radar, often the radar has a (much) larger field of view and a larger range. Also, often radar not only reports on range, but also target velocity and signal strength. As such, this work is more relevant to situations where an imaging system is combined with a laser range finder.

In literature often is reported the use of LADAR (laser radar). Such a LADAR often is implemented as a scanning laser range finder. To the knowledge of the author such systems are not currently available within the Dutch armed forces, and acquisition of such equipment also is not foreseen. As such, this equipment is not chosen as interesting topic of this study.

Section 2 reports what is published in the current literature on this topic. In section 3 an example is given for both a land and sea based scenario showing what can be gained by using range information for automatic target recognition. Finally, conclusions can be found in section 4.

## 2. LITERATURE

The major aim of the literature search was to find publications which report on the combination of imagery and laser range finders to improve object recognition. However, using the available tools for literature search (on-line search in databases containing titles, keywords and abstracts) did not result in any hits reporting on this subject. Consequently, one can say that not very much in this field has been published.

To be able to place the research described in this report within already published work, the literature search is widened to include articles covering very similar topics to the main theme of this report. Note that this involves some arbitrary borders regarding the interest fields covered; for many publications cited in this chapter quite similar publications do exist. However, the work presented here is selected to be representative for the topics covered in literature.

Nandhakumar and Aggarwal [10] give an overview article on the work done on sensor fusion in the computer vision field. Within this field many different types of fusion are discussed, but most types consider fusion of two sensors with the same dimension (e.g. 2D scalar fields) and similar resolution.

<sup>1</sup> E-mail: Schutte@fel.tno.nl, WWW: <http://www.tno.nl>

### **2.1 Combining FLIR with a laser range finder**

Korona and Kokar [9] combine a FLIR and a laser range finder for improved tracking of a landing aircraft. Both sensors are used as input to a Kalman filter. This filter is further improved by using domain knowledge for correction outliers and bias in the results. Romine, Kamen and Sastry [15] describe how to handle the problem (for close / extended objects) that the radar centre of the object might not coincide with the visual centre of the tracked object in a combined Kalman filtering case.

These two publications cover the sensor combination of interest, but the application they discuss is very limited. Also, no obvious way exist to use their work in the field of target recognition.

### **2.2 Combining intensity imagery with dense range data**

Many authors described how to fuse intensity imagery (both visual and IR) with dense range data; often this dense range data is recorded with a laser radar (LADAR.) Chu and Aggarwal [2, 3] present a knowledge based system to interpret laser radar and IR imagery. Results are shown where IR image interpretation is helped with range data. Umeda and Arai [19] describe a system which combines range and intensity imagery to facilitate bin-picking. Harvey and Heinemann [8] describe a system based on the fusion of video and laser radar images or the classification of military ground based targets. The fused results show a drastic increase of performance compared to the single sensor results. Rogers, Tong, Kabrisky and Mills [14] also present a system to segment images using both LADAR and IR imagery. Targets are distinguished from the background by using the fact that the range gradient is smaller for targets than for clutter. Devy and Boumaza [5] propose a system for model-based recognition of objects using both colour and range imagery. Selzer and Gutfinger [16] propose a system which first detects objects in FLIR and LADAR imagery based on edge detection and matching against wire frame models. Bayes or Dempster-Shafer style reasoning fuses the results of the different sensors. Improved results are reported when sensor fusion is used.

Robmann and Bunke [13] present a system to classify edges and junctions based on both range and intensity imagery. Olsson and Gruber [11] describe how to implement an edge closing scheme based on these two sensor modalities.

Wallace, Zhang and Austin [20] describe how to fuse imagery with dense range data. Several levels of fusion are distinguished and described. Similar work is reported by Deguchi [4]. He reports geometrical equations, based on Wu's mechanical theorem proving method, which should be fulfilled to reconstruct 3D target shapes. Pien and Gauch [12] describe to use a variational approach to fuse intensity and range images on pixel level; for this they utilise a Lambertian assumption for the intensity image. Results on synthetic imagery only are presented.

Schwickerath and Beveridge [17] describe a system to determine the registration between LADAR and intensity images.

### **2.3 Conversion towards dense range data**

Tate and Li [18] propose a method to combine high resolution stereo imagery with low resolution range data to produce high resolution range data. The initial range data is combined with a conventional stereo imagery pyramidal scheme, helping ambiguities normally present. Such a scheme could be helpful in transforming laser range finder data, combined with stereo imagery, into dense range data.

### **2.4 Simultaneous visualisation of range and intensity data**

Goss, Beveridge, Stevens and Fuegi [6] propose a method to display of range and intensity imagery simultaneously. It is based on 2.5D display of the intensity imagery. Interaction with a fast workstation is needed for full information retrieval. In [7] they extend this system to also visualise CAD models of objects found with model-based object recognition.

Beveridge, Hanson and Panda [1] describe a complete environment for fusion, ATR and visualisation of the combination of FLIR, colour and LADAR imagery. Basically, this is the combination of the work co-authored by Beveridge [6], [7], [17].

### 3. IMPROVING RECOGNITION USING RANGE INFORMATION

In this chapter for both a land based scenario and an airborne (over sea) scenario is shown that using range information during interpretation of an image results in better performance. This can be expected as range indicates the scale of interest during interpretation, and thus only that scale has to be studied. This should result in a lower false alarm rate without degradation of recognition rate.

#### 3.1 Land based scenario

Research has been performed on a set of images from tape 7A of the CHAR3 trial held at White Sands (USA.) This trial was organised by the NATO group AC243/CCD/WGD in July 1996. It consists a set of IR images recorded from a helicopter depicting a tank. Together with the IR imagery, spoken range information was given on the audio track of the video tapes used.

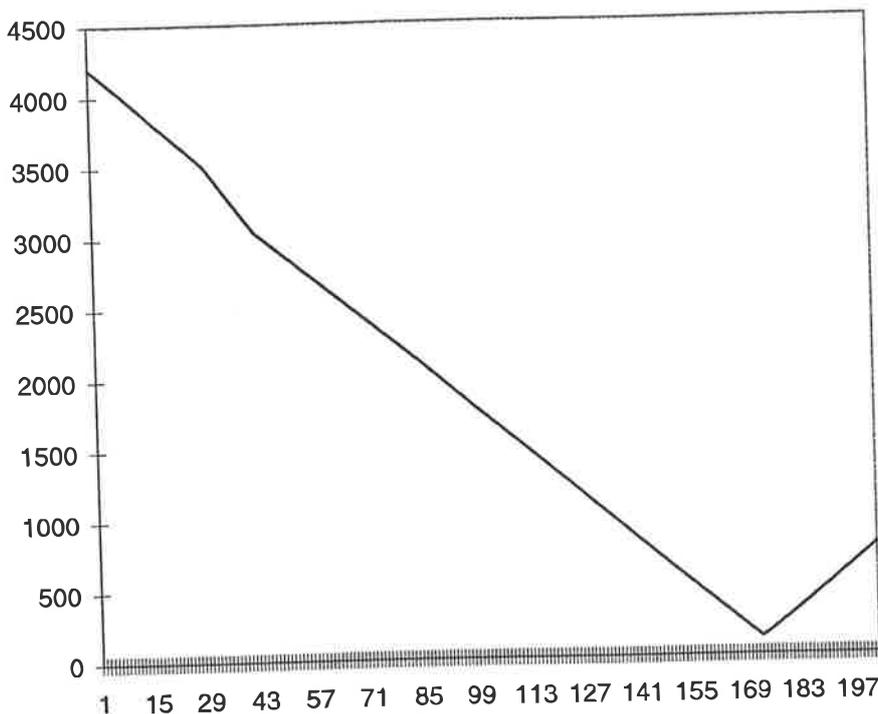


Figure 3.1: Euclidean distance in meters of helicopter to the tank, based on the data spoken on the audio track of the video tape used.

Figure 3.1 shows the range to the tank in meters for each frame in the 203 frame sequence. This graph is only based on a few points, so it claims more accuracy than actually will be present. Figure 3.2 shows an example frame from the sequence. It can be seen that, although the object of interest can be recognised, the task of detection is not entirely trivial. Areas of problems are the stripe character due to the scanning camera, the noise level in a single image frame, and similarity between the characteristics of objects and background. The key difference between background and objects is that hot spots within the objects are surrounded by relatively cold surroundings. To detect this, we are looking for hot spots at a certain scale, compared to its surrounding. When the scale is too large, too much background will be present in the object sample, leading to a too small difference. When the scale is too small, the object influences the estimation of local image statistics too much, leading to a too small estimate of the signal to noise ratio (SNR). One could enlarge the amount of background used when estimating the local image statistics to compensate for this effect, but this will mean that the smooth variations and larger scale edge structure in the background will lead to a larger estimate of the local

variation in the image, and thus to a too small SNR. Concluding, this means that proper scale selection will be essential for good performance.

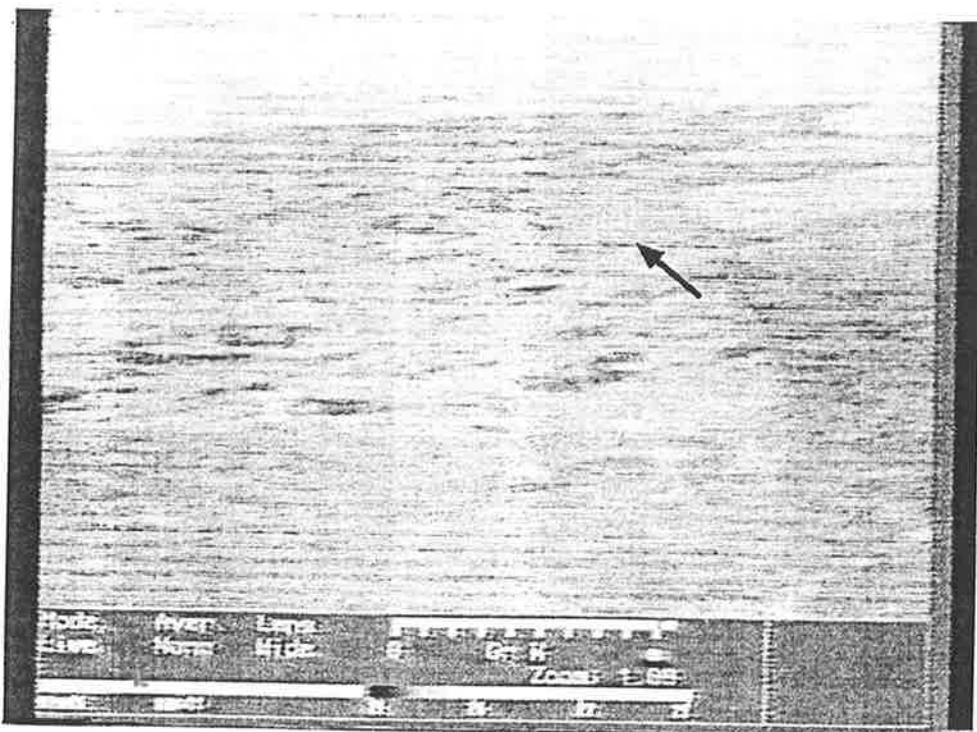
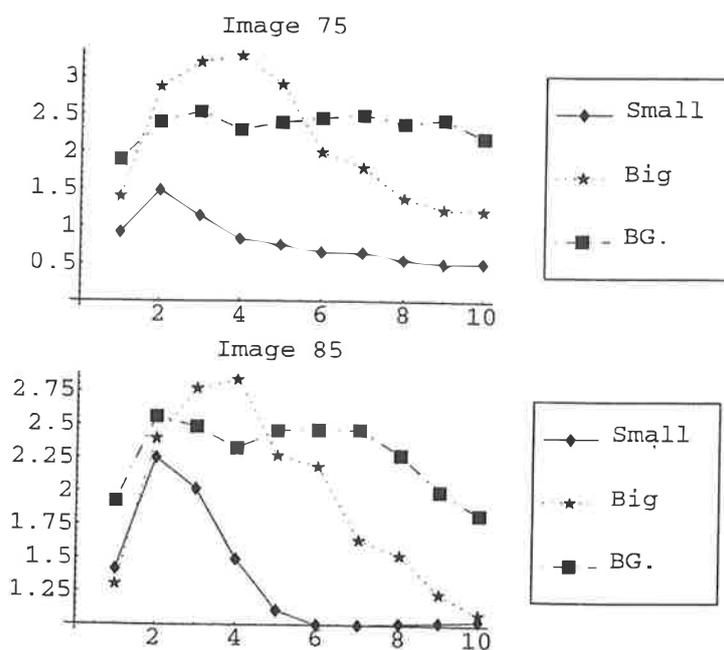


Figure 3.2: Image 100 of the White Sands sequence. The arrow points to the tank to be recognised. Left under the tank a bigger bright spot can be seen corresponding to some buildings in scene. For the experiments described in this report only the imaged area in the centre is used, which has been processed to compensate for different gain of individual detector cells.



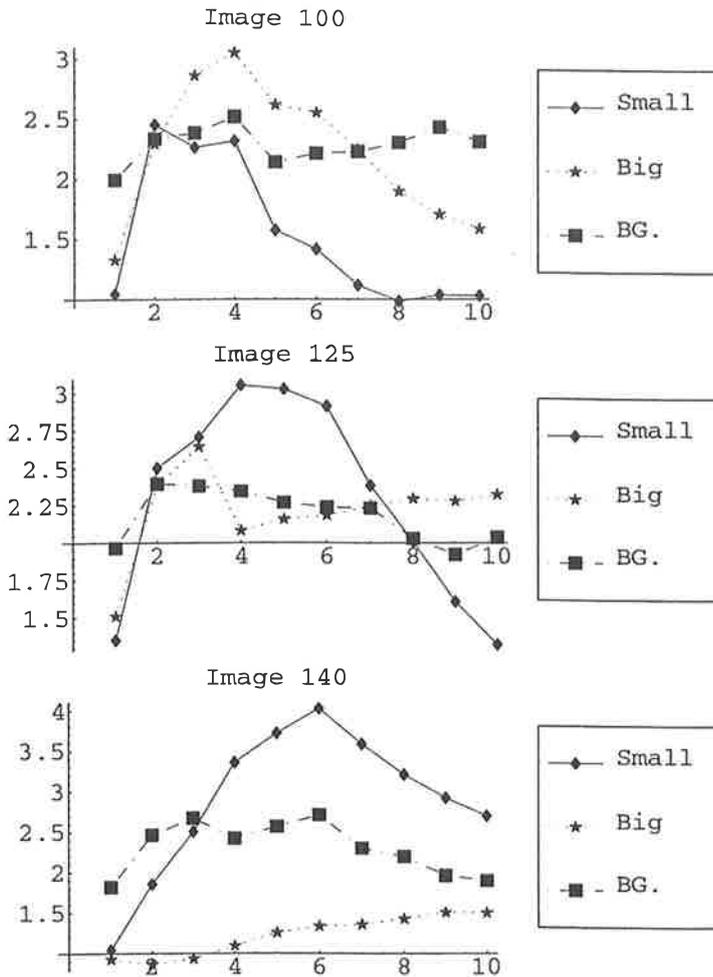


Figure 3.3: Scale dependency of output of detector for image frames 75, 85, 100, 125 and 140. The small blob is the object of interest. The big blob relates to buildings in the scene. Background is the output of the rest of the image. Horizontally is the scale parameter, and vertically the maximal detector output for the different image parts.

Figure 3.3 gives the output of the detection process for some selected frames. The results are not given for all frames, due to problems of getting the processing robust enough to work on a large range of scales. For the small object it can be seen that the optimal scale increases as the distance (as shown in Figure 3.1) decreases. The algorithm used is not sensible for scales smaller than 2, which explains why for larger distances (earlier frames) the signal for the small object is not higher than the background. For smaller distances (later frames), the big object becomes extremely big (it is even closer than the small object to which Figure 3.1 refers). As such, even setting the scale parameter to 10 is not enough to get a

higher output, but it is reasonable to expect that higher settings of the scale parameter will allow detection of such large objects. Within all image frames it can be seen that the detector response for the background is similar, with a maximum response of 2.5 for small scales and slowly decreasing for larger scales. This is rather surprising as during the sequence both flying height and elevation angle are changing.

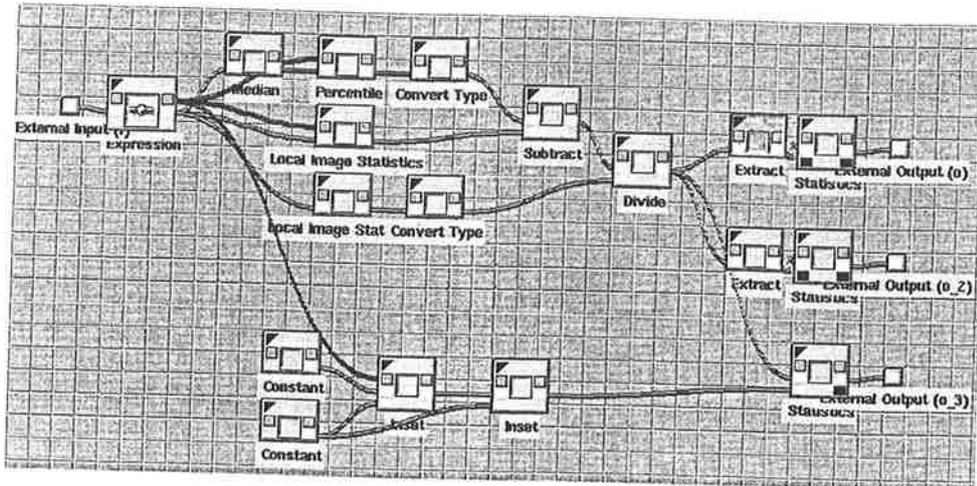


Figure 3.4: Cantata (part of Khoros image processing package) worksheet depicting the processing performed for the landbased scenario. First, the amount of noise is reduced with a median filter. Then, peaks in the image are detected using a percentile (top 25%) filter with a scale-depending size. From this, the local mean is subtracted, and the result is divided by the local standard deviation of the image. These local statistics also are scale dependent. Finally, this result is split into the sets representing the small blob, the big blob and the background for evaluating purposes.

### 3.1.1 Conclusions

From the results it can be deduced that the optimal scale parameter for object detection is related to the distance to the object. As such, using range information within the object detection process will improve the detection rate.

## 3.2 Airborne scenario

Parts of the work in this section is performed within the framework of EUCLID RTP-8.2 "Intelligent Sensors" as part of the START consortium. As such, it is published in deliverable WE/114.3F, Trade-off between different performance improvements due to fusion data deriving from non-infrared sensor. Within that work, our collaboration and discussion with SAGEM is acknowledged.

The problem at hand is the detection of a propeller plane which is flying away from the observer. This plane is observed in a sequence of 100 frames, example frames are found in Figure 3.5.

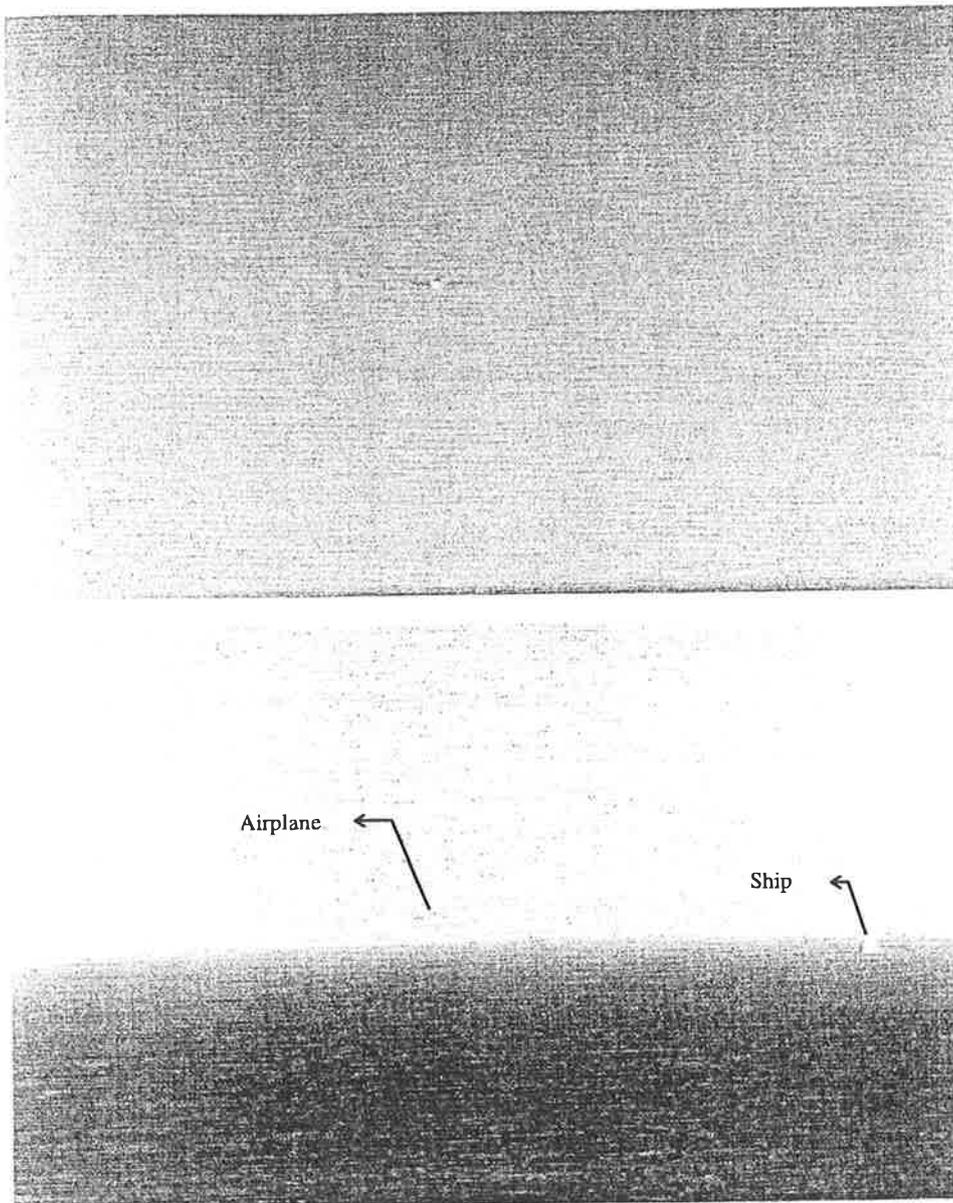


Figure 3.5: frame 11 (top) and frame 99 (bottom) of the sequence. In the bottom image the right object is a ship which is not an object class of interest.

The range information used is synthetic. The information is calculated using a known speed  $v$  of the plane (90 knots  $\equiv$  45m/s), the known frame rate  $r$  (25/13 Hz), the known IFOV  $i$  of the camera used (0.12 mrad), the measured size of the plane  $p$  in frame 11 (43 pixels from wingtip to wingtip) and its estimated size  $w$  (7 meter). The associated formula for the range  $r$  is for frame  $n$ :

$$r = \frac{w}{i \cdot p} + \frac{n - 11}{f} \cdot v$$

which equates to 1.0 km for frame 0, 1.3 km for frame 11 and 3.4 km for frame 99.

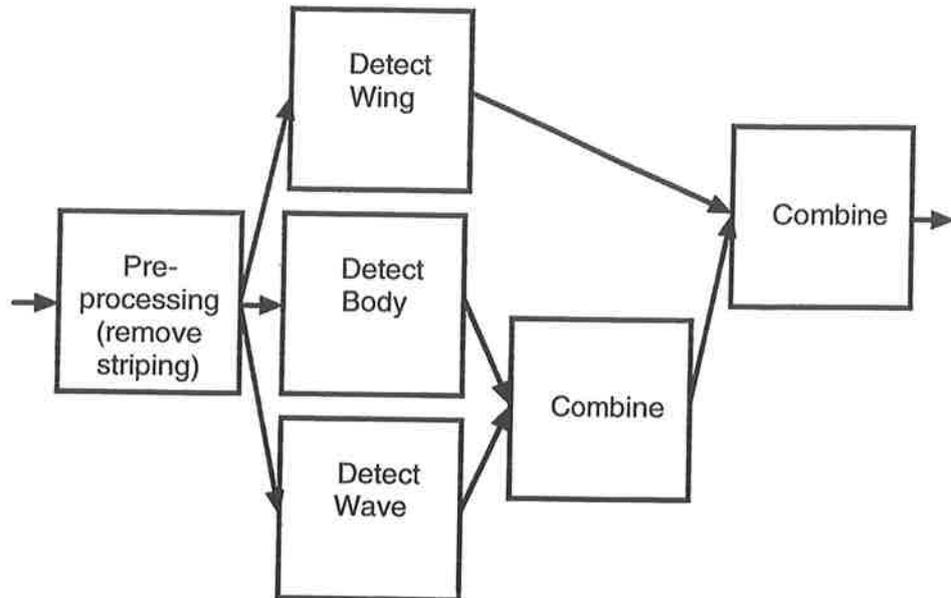


Figure 3.6: Schematic overview of the object detection used. First, artefacts (striping) due to different gain of different sensor elements are reduced by subtracting the per-line average. After that, using morphological operations three different classes of objects are detected: wings, aeroplane bodies and waves. The size of the structuring element is linked to the scaling parameter  $s$ . The body detector is also sensitive to waves at sea; in the combine of the wave and body detector the detected waves are cancelled out. The combine step of that and the wing detector seeks in the neighbourhood of the body for a wing; if that is not found that body is cancelled.



Figure 3.7: Model of the plane used: a circular body, hot compared to its surroundings, is in the vicinity of two relatively cold, elongated wings.

The algorithm used is described in Figure 3.6. It consists of three logical parts: wing detection, body detection and clutter rejection. The wing and the body detector are sensitive to scale: it is very hard to make a detector which (with an equally low false rate) is sensitive over large variations in scale. The wing detector is based upon a grey level dilation with as structuring element the shape of a wing, which is scaled to local image mean and variance. The body detector is based on the difference between the mean of a small area with its local mean. This should exceed a threshold value, and also this value scaled to local variance should exceed a threshold value. However, the body detector still is sensitive to large bright objects such as the waves in the sea. For that reason a clutter rejection filter is added, which detects waves as large bright blobs using the grey level erosion with a fairly large structuring element (i.e. the model for the wave) scaled to local mean and variance. At regions where near a body a wing is found we conclude that an aeroplane is present. All together we have several places where notions like small, large and near are used -- which are all keyed to the apparent size of the object. As the range of the object is different through the scene this apparent size, called scale in the rest of this document, is changing due to the perspective projection inherent to an IR imager. Using known range information, we can predict for each frame the optimal scale for the algorithm.

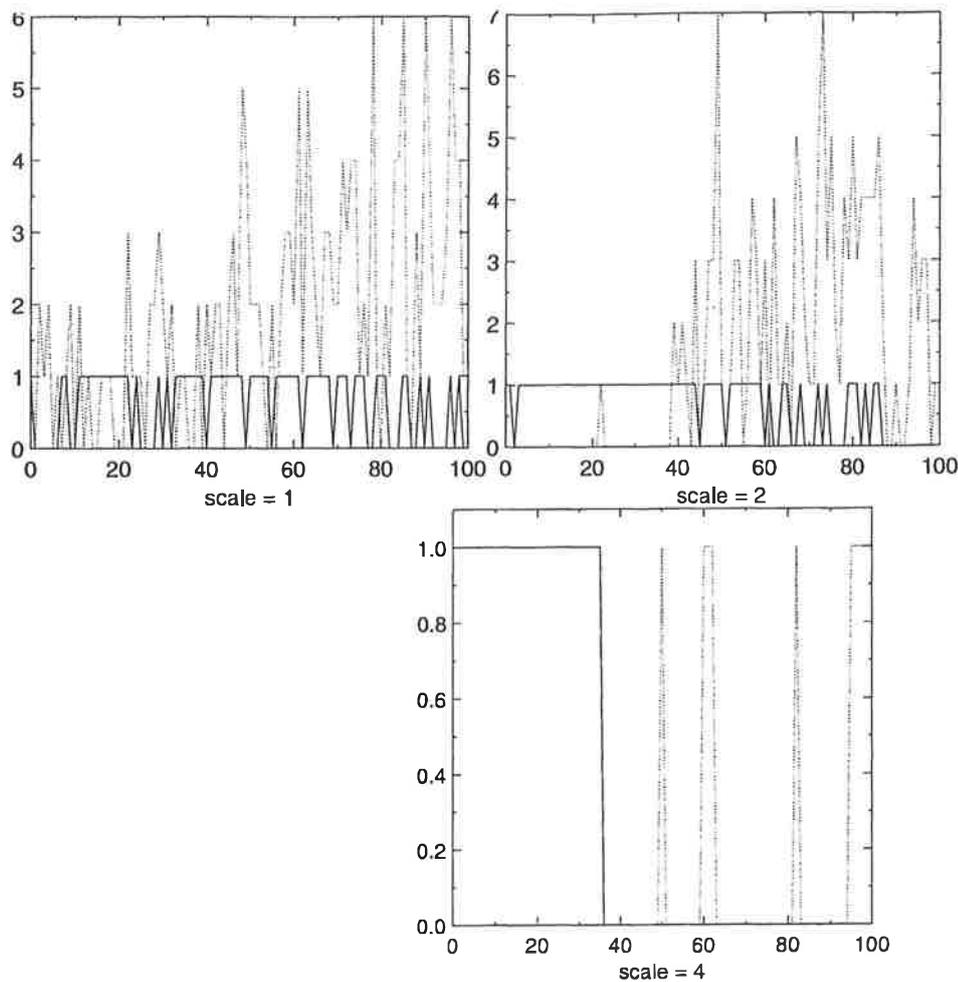


Figure 3.8: The detection performance of the algorithm for the scale parameter = 1, 2 and 4. The solid line represents the number of correct detections per frame, the dotted line the number of false alarms per frame.

In Figure 3.8 can be seen that at scale = 1 we have reasonable detection performance between frame 30 and frame 99. At scale = 2 we have good detection performance up to frame 60. At scale = 4 we have excellent detection performance up to frame 38. A standard multi-scale approach will add the detection results from the different scales. This results, as shown in Figure 3.9 in a good detection performance over the whole range. However, this also means that we have the false alarms generated at each scale parameter, witnessed by the rather high false alarm rate of this combination. More advantageous is to select the correct scale based on the range information. Here we use scale = 4 for distances up to 1100 meter, scale = 2 for distances between 1100 and 2200 meter, and scale = 1 for distances over 2200 meter (this corresponds to frame 0-2 for scale = 4, frame 3-50 for scale = 2, and frame 51-99 for scale = 1.)

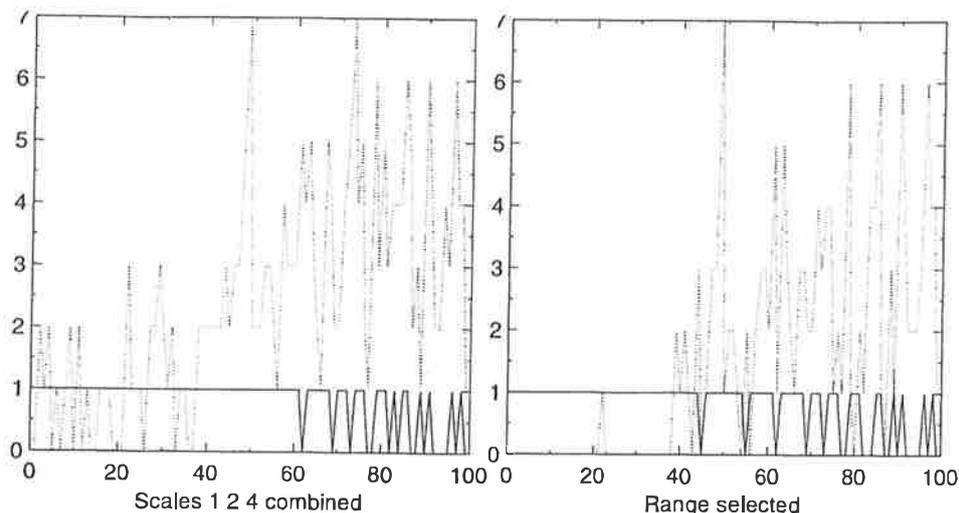


Figure 3.9: left: The performance for the combination (adding) of scales 1, 2 and 4. right: The performance for the range selection between scale 1, 2 and 4. For the combination of scales that false alarm rate is used which is the maximum found in one of the scales, as it can be expected that all scales denote the same false alarms (and thus addition of false alarms will count the same object multiple times.)

Table 3.1 gives the numeric results which are depicted in Figure 3.8 and Figure 3.9. Also it shows the results if these detection results are used as input to a tracker. The tracker used is a simple nearest neighbour tracker with a rectangular gate. A track is confirmed using the n out of 5 rule, with results for n = 2 and n = 3 shown. At n = 3 none of the tracking schemes gave raise to a false alarm. It can be seen that the range selected tracker is superior to the other four trackers; at n = 2 a detection performance of 95% is reached without any false alarms. The lower correct detection percentage of the combined scale tracker at n = 2 probably is due to the fact that the plots of the three scales are updated as a combined plot in the tracker, which can lead to a lot of track breaks and crossovers. It can be expected that with the proper logic this detection percentage can be at least as high as that with the selected range tracker.

Table 3.1: Detection performance for different scales and combinations of scales. Combined means that the detections of scale 1, 2 and 4 are added. Selected means that for frame 0 to 2 scale 4 is used, for frame 3 to 50 scale 2 is used and for frame 51 to 99 scale 1 is used. Correct detections is the sum of the number of detections per frame near the correct location. False detections is the sum of number of false alarms per frame. False tracks is the number of erroneous tracks at track threshold = 2 in the whole sequence; False track fr. is the sum over all frames of the number of false tracks at track threshold = 2 per frame. Correct track n=2 and Correct track n=3 is the number of frames with tracks near the correct location at track threshold 2 respectively 3.

	Correct detections	False detections	False tracks	False track fr.	Correct track n=2	Correct track n=3
Scale = 1	65	187	2	4	84	48
Scale = 2	69	141	3	5	79	62
Scale = 4	36	10	1	2	38	33
Combined	84	241	7	16	92	70
Selected	81	151	0	0	95	64

### 3.2.1 Conclusions

This example shows how range information can be incorporated in a processing algorithm. Compared to single scale solutions it can boost detection rate from 69% to 81%, with a minimal increase in false alarms. When combined with a tracker the results are even more spectacular: the detection rate is increased from 84% to 95%, while the two erroneous

tracks previously has disappeared. Note however that these results are obtained by processing a single sequence, with an algorithm and parameters tuned specially for that sequence.

#### 4. DISCUSSION

A very remarkable result of this study that very little work is published in the field of fusion between sparse range data and imagery. This could be expected when little advantage is gained by this combination, but the results in chapter 3 indicate the contrary. An explanation for this could be that many researchers in the past in the field of image automatic interpretation aimed at the most general object detection algorithm possible, which includes scale invariance. Within the approach proposed in this report the exact opposite direction is taken: introduce scale dependence, know the correct scale and use that to your advantage.

The results given in chapter 3 show that the fusion of range information with imagery, especially in combination with tracking, gives a significant increase in the performance of the system in terms of detected object / false alarm ratio.

Although the use of LADAR was not the topic of this study, the literature search revealed that such equipment certainly is interesting. As such, a study towards this kind of sensors (including the fusion with the intensity channel) seems worthwhile.

#### 5. REFERENCES

- [1] J.R. Beveridge, A. Hanson and D. Panda, *Model-based Fusion of FLIR, Color and LADAR*, Sensor Fusion and Networked Robotics VIII, SPIE Vol. 2589, pp. 2-11, 1995
- [2] C.C. Chu and J. K. Aggarwal, *Multi-Sensor Image Interpretation Using LADAR Radar and Thermal Images*, Interim technical report, TR-6-70, Computer and Vision Research Center, The University of Texas at Austin, 1991
- [3] C.C. Chu and J. K. Aggarwal, *Image Interpretation Using Multiple Sensing Modalities*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 14, pp. 840-847, 1992
- [4] K. Deguchi, *An Algebraic Framework for Fusing Geometric Constraints of Vision and Range Sensor Data*, Proc. of the 1994 IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems (MFI'94), Las Vegas, pp. 329-336, 1994
- [5] M. Devy and R. Boumaza, *Multi-Sensory Fusion and Model-Based Recognition of Complex Objects*, Proc. of the 1994 IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems (MFI'94), Las Vegas, pp. 345-352, 1994
- [6] M.E. Goss, J.R. Beveridge, M. Stevens and A. Fuegi, *Visualization and Verification of Automatic Target Recognition Results Using Combined Range and Optical Imager*, Proceedings: Image Understanding Workshop, ARPA, pages 491-494, 1994
- [7] M.E. Goss, J.R. Beveridge, M. Stevens and A. Fuegi, *Three-dimensional visualization environment for multisensor data analysis, interpretation, and model-based object recognition*, Visual Data Exploration and Analysis II, SPIE Vol. 2410, pp. 283-291, 1995
- [8] R. L. Harvey and K. G. Heinemann, *Biological vision models for sensor fusion*, IEEE Conference on Control Applications, pp. 392-397, 1992
- [9] Z. Korona and M.M. Kokar, *Landing aircraft tracking using fusion of FLIR and laser range finder*, Sensor Fusion and Aerospace Applications, SPIE Vol. 1956, pp. 139-150, 1993
- [10] N. Nandhakumar and J.K. Aggarwal, *Multisensory Computer Vision*, Advances in Computers, Vol. 34, pp. 59-111, 1992
- [11] J. Olsson and S. Gruber, *Fusion of real and logical sensors derived from range and intensity images*, Sensor Fusion and Aerospace Applications, SPIE Vol. 1956, pp. 187-198, 1993
- [12] H.H. Pien and J.M. Gauch, *A Variational Approach to Multi-Sensor Fusion of Images*, Applied Intelligence, Vol. 5, pp. 217-235, 1995
- [13] R. Robmann and H. Bunke, *Towards Robust Edge Extraction - A Fusion Based Approach Using Greylevel and Range Images*, Automatic Object Recognition V, SPIE Vol. 2485, pp. 249-259, 1995
- [14] S. K. Rogers, C. W. Tong, M. Kabrisky and J. P. Mills, *Multisensor fusion of LADAR and passive infrared imagery for target recognition*, Optical Engineering, Vol. 28, No. 8, pp. 881-885, 1989

- [15] J.B. Romine, E.W. Kamen and C.R. Sastry, *Fusion of radar and imaging sensor data for target tracking*, Signal and Data Processing of Small Targets, SPIE Vol. 2235, pp. 558-569, 1994
- [16] F. Selzer and D. Gutfinger, *LADAR and FLIR based sensor fusion for automatic target classification*, Sensor Fusion: Spatial Reasoning and Scene Interpretation, SPIE Vol. 1003, pp. 236-246, 1988
- [17] A. N. A. Schwickerath and J. R. Beveridge, *Object to Multisensor Coregistration with Eight Degrees of Freedom*, Proc. Image Understanding Workshop, Monterey, CA, pp. 481-490, 1994
- [18] K. Tate and Z-N. Li, *Depth Map Construction from Range-Guided Multiresolution Stereo Matching*, IEEE Trans. on Systems, Man, and Cybernetics, Vol. 24, pp. 134-144, 1994
- [19] K. Umeda and T. Arai, *Industrial Vision System by Fusing Range Image and Intensity Image*, Proc. of the 1994 IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems (MFI'94), Las Vegas, pp. 337-344, 1994
- [20] A. Wallace, G. Zhang and B. Austin, *Fusion of intensity and range data*, IEE Colloquium on '3D Imaging and Analysis of Depth/Range Images', London, pp. 7/1-10, 1994