Long-Range Target Detection in a Cluttered Environment using Multi-Sensor Image Sequences

D. Borghys, P. Verlinde, C. Perneel and M. Acheroy

Royal Military Academy, Signal & Image Centre,
Renaissancelaan 30, 1000 Brussels, Belgium

ABSTRACT

An approach to the long range automatic detection of vehicles, using multi-sensor image sequences, is presented. The algorithm was tested on a database of six sequences, acquired under diverse operational conditions. The vehicles in the sequences can be either moving or stationary. The sensors also can be moving. The presented approach consists of two parts.

The first part detects targets in single images using seven texture measurements. The values of some of the textural features at a target position will differ from those found in the background. To perform a first classification between target- and non-target pixels, linear discriminant analysis is used on one test image for each type of sensor. Because the features are closely linked to the physical properties of the sensors, the discriminant function also gives good results to the remainder of the database sequences. By applying the discriminant function to the feature space of textural parameters, a new image is created. The local maxima of this image correspond to probable target positions. To reduce the false alarm rate, any available prior knowledge about possible target size and aspect ratio is incorporated using a region growing procedure around the local maxima.

The second part of the algorithm detects moving targets. First any motion of the sensor itself needs to be detected. The detection is based on a comparison of the spatial cooccurrence matrix within one image and the temporal cooccurrence matrix between successive images. If sensor motion is detected, it is estimated using a multi-resolution Markov Random Field (MRF) model. Available prior knowledge about the sensor motion is used to simplify the motion estimation. The motion estimate is used to warp past images onto the current one. Moving targets are detected by thresholding the difference between the original and warped images. Temporal and spatial consistency are used to reduce false alarm rate.

Final results are obtained by fusing the results for the different sensors and the two parts of the algorithm.

Keywords: Sensor Fusion, Target Detection, Texture, Motion Estimation

1. INTRODUCTION

Long range automatic detection of vehicles is of great military importance to modern armed forces. The most critical factor of any system for automatic detection is its ability to find an acceptable compromise between the probability of detection and the number of false targets. A lot of work has already been carried out on the detection of single vehicles and target formations.[1–3] However, detection and tracking of small, low contrast vehicles in a highly cluttered environment using a single sensor, still remains a very difficult task.

This paper describes an approach to tackle this problem using sensor fusion. The approach was implemented and then tested on a set of six image sequences obtained from different sensors under diverse operational circumstances. The target area in the images varies from small, typically less than 10 pixels-on-target, up to 300 pixels-on-target. The vehicles can be either moving or stationary. In most of the sequences, the sensor was mounted on a stationary platform and could only perform a tilt and pan operation. Most of the images are highly cluttered. This clutter is caused by sensor noise, natural background texture and the presence of human artifacts in the scene (e.g. buildings).

The approach presented in this paper consists of two independent parts for each sensor. The next section presents a global overview of the method. The subsequent sections respectively describe the two parts of the algorithm and the fusion step. The last two sections show results and conclusions.

Other author information: Send correspondence to D.B.: Email: dirk@elec.rma.ac.be; Tel: 32-2-737.61.61; Fax: 32-2-737.61.63

2. OVERVIEW OF THE APPROACH

For each sensor the proposed algorithm consists of two independent parts. The first part searches for targets in single images while the second part uses multiple subsequent images in order to specifically find moving targets. For each sensor, each part of the algorithm behaves as an expert indicating the possible presence of vehicles in the scene. The outcomes from all experts are fused to reach a final decision. Figure 1 shows the global overview of the method.

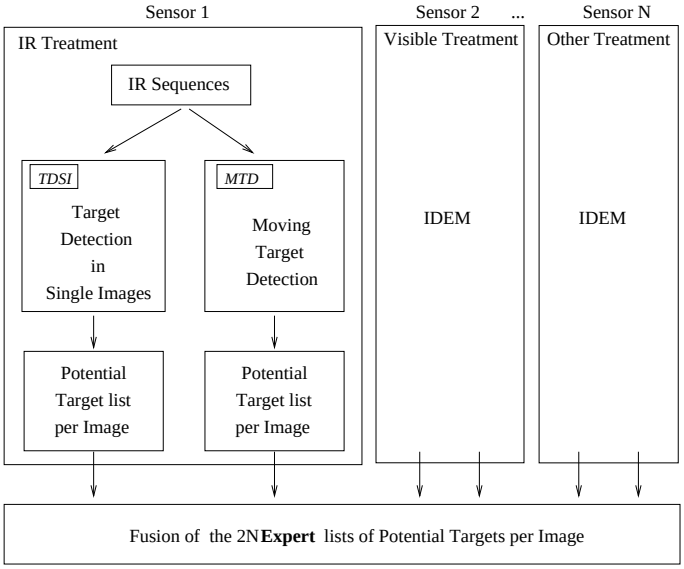

Figure 1. Overview of the approach

3. TARGET DETECTION IN SINGLE IMAGES (TDSI MODULE)

3.1. Introduction

For the detection of targets in single images, a semi-supervised approach based on texture measurements was chosen. For each sensor type, one image was selected to constitute the learning database. On these images the true targets were delimited. Then several texture parameters were calculated at each pixel of these test images and linear discriminant analysis was used to find a linear combination of the texture parameters that optimally discriminated the targets from the background.

The actual detection algorithm then applies the same discriminant function to the texture measurements calculated on the remainder of the image database. When this function is applied at each pixel of an image, a new image can be formed in which the maxima correspond to likely target positions. To find the possible target positions, first the local maxima are determined in this new image and then available prior knowledge about possible target size and aspect ratio is used to reject false targets in single images.

3.2. Texture parameters

The calculation of the texture measurements is based on the cooccurrence matrix. The cooccurrence matrix is defined as a function of a given direction and distance, or alternatively, as a function of a displacement (dx,dy) along the x and y direction in the image. For a given displacement (dx,dy), the (i, j) element of the cooccurrence matrix is the number of times the grey value at the current position (x,y) is i when the value at the distant position (x+dx,y+dy) is j.

$$C^{dx,dy}(i,j) = P(G(x,y) = i \mid G(x+dx, y+dy) = j) \tag{1}$$

The cooccurrence matrix can be calculated on the whole image. However, by calculating it in a small window scanning the image, a cooccurrence matrix can be associated with each image position. The centre of the window is

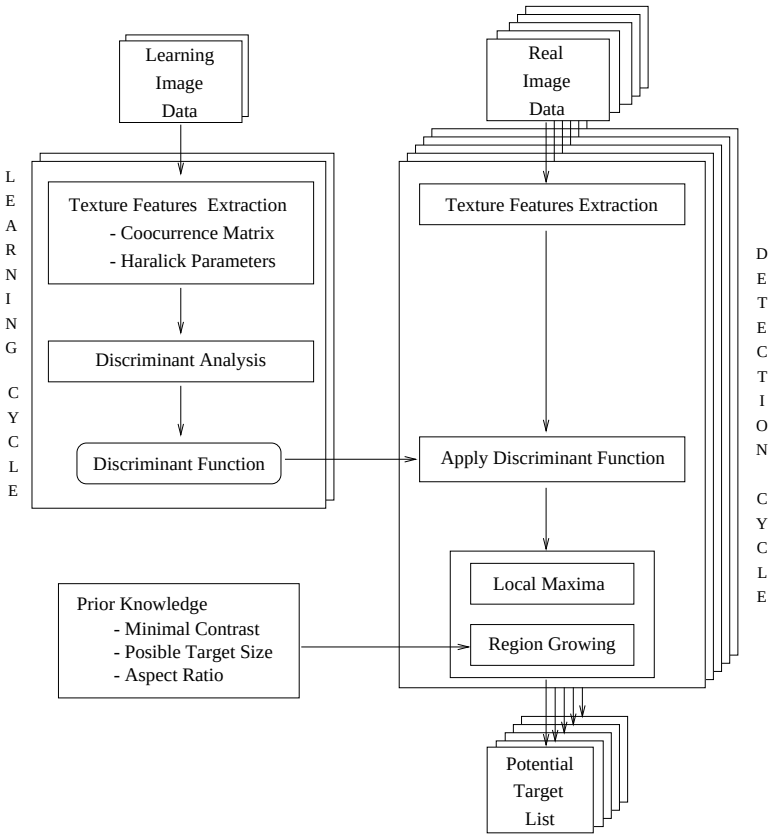

Figure 2. TDSI module

denoted (x_c, y_c) and the corresponding cooccurrence matrix is $C^{dx,dy}_{x_c,y_c}(i,j)$ In figure 3 an example of a cooccurrence matrix is shown. The matrix corresponds to the small window of the image on the left and was calculated for a displacement of $dx = 1, dy = 2$.

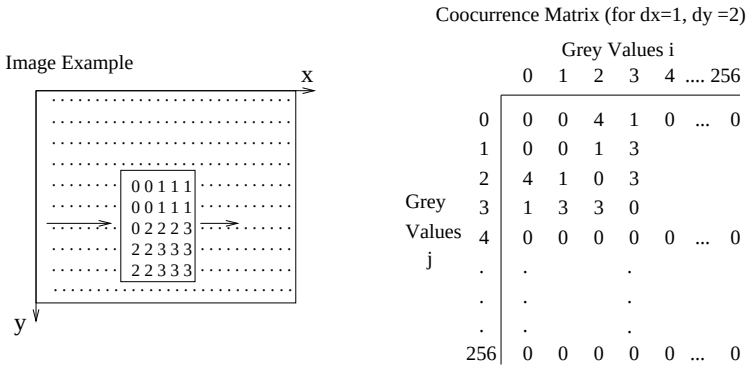

Figure 3. Cooccurrence matrix

The textural features that were used, were introduced by Haralick[4–6] and are widely used in texture analysis.

Based on the local cooccurrence matrix, the used parameters are defined as follows:

$$
\begin{aligned}
F_1(x_c,y_c) = \quad & Energy && = \textstyle\sum_i \sum_j C^{dx,dy}_{x_c,y_c}(i,j)^2 \\
F_2(x_c,y_c) = \quad & Contrast && = \textstyle\sum_i \sum_j \left[(i-j)^2 C^{11}_{x_c,y_c}(dx,dy)\right] \\
F_3(x_c,y_c) = \quad & Max.Prob. && = max\ \left[C^{dx,dy}_{x_c,y_c}(i,j)\right] \\
F_4(x_c,y_c) = \quad & Entropy && = C^{dx,dy}_{x_c,y_c}(i,j) log\left[C^{dx,dy}_{x_c,y_c}(i,j)\right] \\
F_5(x_c,y_c) = \quad & Homogeneity && = \textstyle\sum_i \sum_j \frac{max\left[C^{dx,dy}_{x_c,y_c}(i,j)\right]}{[1+(i-j)^2]} \\
F_6(x_c,y_c) = \quad & Variance && = \left[\textstyle\sum_i (i-E_i)^2 \sum_j C^{dx,dy}_{x_c,y_c}(i,j)\right]\left[\textstyle\sum_j (j-E_j)^2 \sum_i C^{dx,dy}_{x_c,y_c}(i,j)\right] \\
& && with\ E_i = \textstyle\sum_i i \sum_j C^{dx,dy}_{x_c,y_c}(i,j) \\
& && and\ E_j = \textstyle\sum_j j \sum_i C^{dx,dy}_{x_c,y_c}(i,j)
\end{aligned}
\tag{2}
$$

Because we are not interested in modelling texture but only want in detecting a difference between target and background pixels, an arbitrary displacement of $dx = 1, dy = 1$ was chosen for all calculations. The window used to calculate the local cooccurrence matrix had a size of 5×5. The results for each texture feature can be converted into an image. Figure 4 shows an example of an infrared and a visual image and figure 5 shows the corresponding texture images.

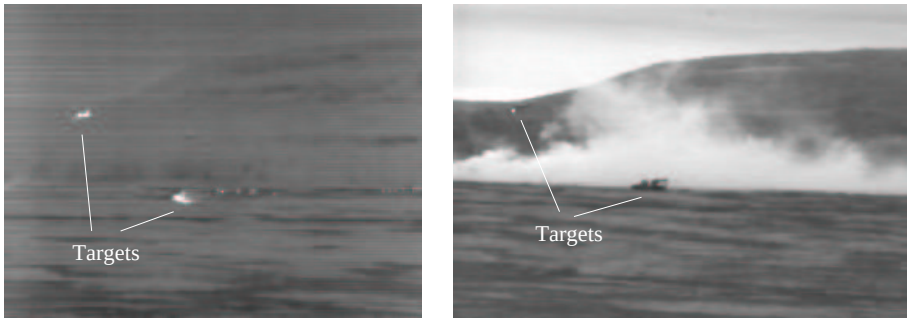

Figure 4. Example of IR and Visual image (Courtesy of Defense Research Establishment Valcartier, Quebec, Canada)

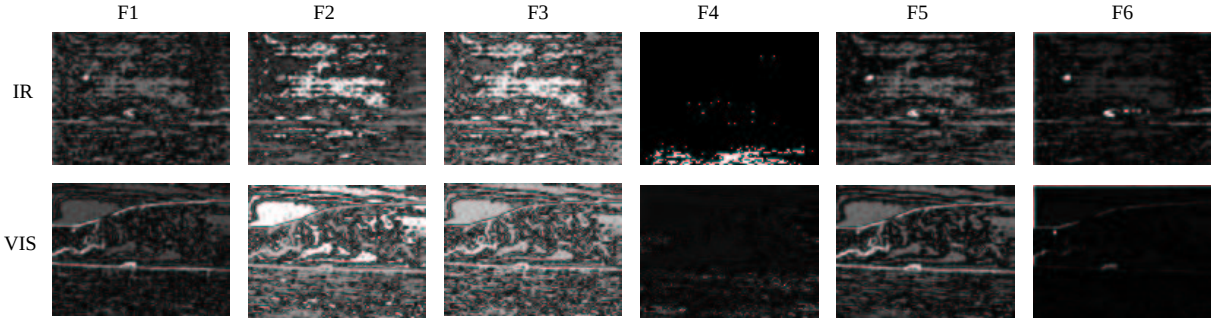

Figure 5. Texture Images

As can be noticed in figure 5, the vehicles appear very clearly in some of the texture images. Hence the idea to combine these features to get an optimal discrimination between background and targets. In order to achieve this, linear discriminant analysis was applied to the texture images of the learning set of images. We used Wilks's stepwise method[7] to find the coefficients W(i) of the linear discriminant function:

$$V(x,y) = \sum_i F_i(x,y)W(i) \tag{3}$$

where $V(x,y)$ is maximised for a target pixel and minimised for a background pixel.

Table 1 shows the results for the visual and infrared images. Note that, for infrared images, contrast, variance and homogeneity are the most discriminating features. This corresponds to the intuitive idea that in infrared images vehicles are seen as hot spots.

Coefficient	Feature	IR type sensors	VIS type sensors
W(1)	Energy	-0.408	0
W(2)	Contrast	1.587	1.563
W(3)	Max. Prob.	0.299	-0.835
W(4)	Entropy	-0.743	-0.604
W(5)	Homogeneity	1.064	1.068
W(6)	Variance	2.539	0

Table 1. Coefficients for discriminant function

3.3. Region Growing around Local Maxima

By calculating the discriminant function $V(x,y)$ in each point of the original image, a new image can be created in which the pixel value is proportional to the probability that the given point belongs to a vehicle. To detect the targets it is thus necessary to find the local maxima in this new image. A region growing procedure around these maxima is then used to incorporate prior knowledge about target size and aspect ratio.

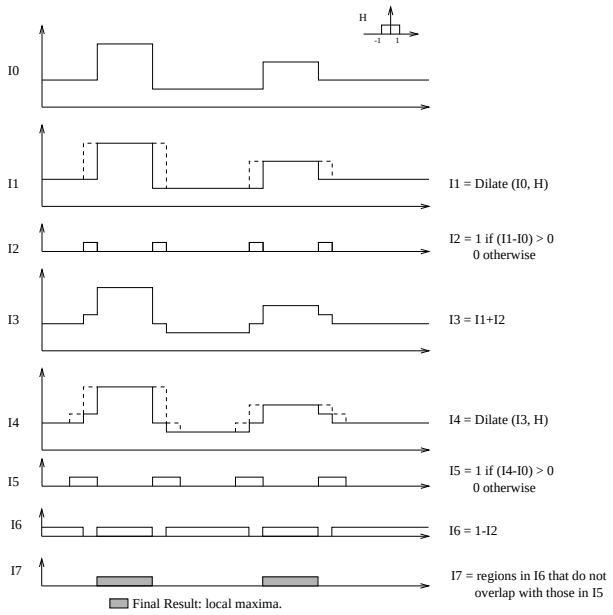

Figure 6. Detection of Local Maxima

3.3.1. Local Maxima

The detection of local maxima is based on a succession of morphological operations.[8,9] The basic operator is a dilation with a 2×2 structuring element. Figure 6 shows the different steps of the method.

3.3.2. Region Growing

To incorporate any available prior knowledge about the possible range of target size or aspect ratio, a region growing procedure is used. The initial regions for the region growing are the local maxima in the image. Surrounding pixels are added to these regions as long as their grey level differs less than a given threshold from the value at the local maxima. If the region becomes too large it is discarded. If the region growing of a given region stops before it reaches the upper size-limit, the other constraints are checked. If a constraint is not satisfied, the region is discarded.

4. MOVING TARGET DETECTION (MTD MODULE)

The second part of the algorithm focusses on the detection of moving targets. In order to detect moving targets, any sensor motion needs to be detected and its effects compensated first. Then, preceding images can be warped onto the current one, and moving objects will appear as a difference between the original image and the warped ones.

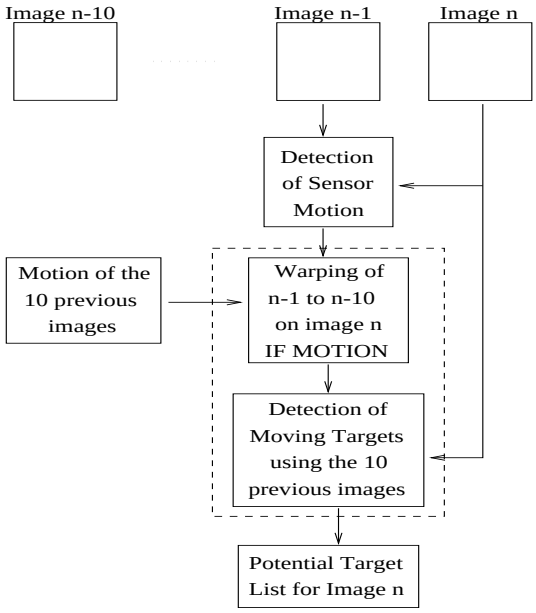

Figure 7. MTD module

4.1. Detection of sensor motion

The detection of sensor motion is again based on cooccurrence matrices. This time the cooccurrence matrix is calculated between an image and the preceding one (temporal cooccurrence matrix).

$$C_{x_c,y_c}^{dx,dy,dt}(i,j) = P(G(x,y;\ t) = i \mid G(x+dx, y+dy;\ t+dt) = j)$$

If no sensor motion occured between the two images, ideally, for $dx = dy = 0$ (i.e. no spatial displacement), all non-zero elements of the temporal cooccurrence matrix should lie on the diagonal. However, due to noise, there will be a small spread along the diagonal. If one calculates the spatial cooccurrence matrix for a small displacement, the spread along the diagonal is due to noise and to the fact that the image is not homogeneous. Therefore, when comparing this spatial cooccurrence matrix with the temporal cooccurrence matrix, the spread along the diagonal is expected to be the largest in the former one if no motion occured between the two images that were used to

calculate the temporal cooccurrence matrix. When motion is present, the spread along the diagonal quickly becomes larger. The measurement we used to detect sensor motion is based on the percentage of off-diagonal points in both cooccurrence matrices:

$$MC = \frac{\sum_j \sum_{j \neq i} C(i,j)}{\sum_j \sum_i C(i,j)}$$

This is calculated for both the temporal (MC_{temp}) and for the spatial cooccurrence matrix (MC_{spat}). Sensor motion is said to be present if $\frac{MC_{temp} - MC_{spat}}{MC_{spat}} \geq 0.005$. In figure 8 the spatial and temporal cooccurence matrix are shown. The upper images show the matrices for a part of a sequence where no sensor motion was present. The lower images show an example of both matrices calculated in a part of the same sequence where the sensor was moving.

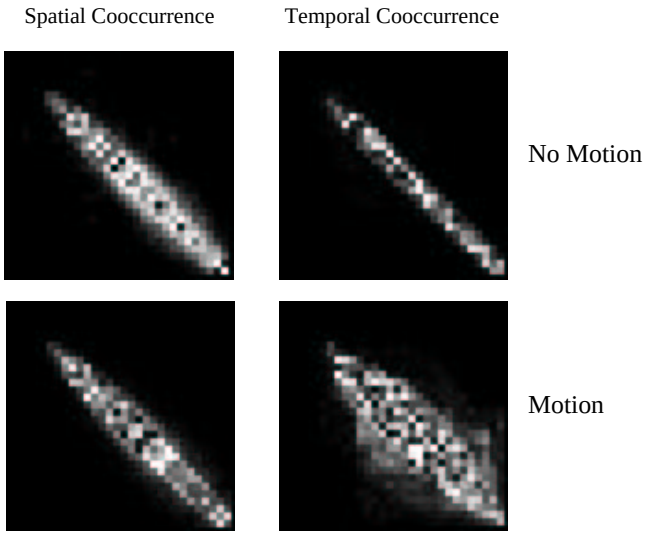

Figure 8. Detection of sensor motion

4.2. Motion Estimation

If sensor motion is detected, we need to estimate it and compensate its effects on the images. The estimation of sensor motion relies on prior knowledge. Three cases are distinguished.

- If it is known that the sensor was fixed and can only do a pan or tilt, the corresponding motion in the image will consist of a uniform translation. In this case we search for the translation by optimising the correlation for a few horizontal and vertical lines.
- If the sensor was mounted in an aircraft, the terrain can be approximated by a plane and a model of the perspective projection of a rigid plane moving in three dimensions is used.[10] This model has only 8 parameters, so finding 4 displacement vectors between two images is sufficient to determine these parameters. In fact we use a threshold that is progressively lowered until at least 15 corresponding regions are found and then a least square method is used to find the parameters.[3]
- In any other case we calculate the optical flow. The calculation of optical flow is based on a Markov Random Field model[11,12] which implements a conservation and a smoothness constraint but at the same time handles motion discontinuities and occlusions.

4.3. Detection of moving targets

Once the sensor motion is estimated, preceding images are warped onto the current one. Then the original image is subtracted from the warped ones. If a moving object is present in the scene, we should find a large value at it position. The resulting images after subtraction are therefore thresholded and objects with acceptable size and aspect ratio are selected using a region growing procedure. Tracking is used to get the target list. Figure 9 shows the result of subtracting the original image from the warped ones.

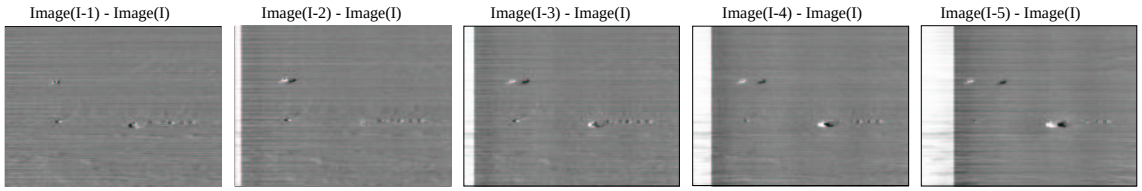

Figure 9. Detection of moving targets

5. DECISION FUSION

Each of the two parts of the algorithm for each of the available sensors behaves as an expert indicating the possible position of targets in the scene. The final decision is reached by fusing the results of these experts. Because, in the current algorithm, each expert only provides a binary decision, the decision fusion is limited to a "k out of N" voting-rule,[13,14] using the results of the different experts.

6. Results and Discussion

Table 2 shows the results for the two parts of the algorithm and this for each sensor. In the table Pd is the probability of detection and Nft is the average number of false targets per image.

Sequence	Sensor	Results of TDSI		Results of MTD	
		Pd	Nft	Pd	Nft
1	LW	93	2	92	1
	VIS	29	4	1	0
2	LW	44	0	20	0
	VIS	89	7	36	0
3	LW	78	4	9	1
	VIS	14	7	7	0
4	LW	87	1	4	0
	SW	0	3	0	0
	VIS	93	0	4	0
5	LW	5	1	4	0
	SW	12	4	3	0
	VIS	98	0	3	0
6	LW	20	1	1	3
	Red	97	0	42	0
	Green	95	0	41	0
	Blue	95	0	46	0

Table 2. Results of the different "experts"

From table 2 it appears that the visual and infrared (i.e. LW or SW) sensor are complementary. In sequences 1 and 3, the infrared sensor gives the best results while in some sequences (2,5,6), it is outperformed by the visual

Sequence	Results for k=1 (OR rule)		Results for k=2	
	Pd	Nft	Pd	Nft
1	97	9	69	1
2	63	11	67	2
3	94	14	77	4
4	98	6	57	0.5
5	98	7	56	1
6	94	7	87	1

Table 3. Results after decision fusion with "k out of N" voting-rule

sensor. The motion detection only gives useful results in the first and sixth sequence. This is basicly due to the sequences themselves. In sequences 2 and 3 the target is approaching the sensor straight on and therefore it is only detected as a moving target at the end of the sequence where its apparent size increases. In sequence 4 and 5 the targets are stationary. The few detections that are made by the MTD algorithm are due to a moving antenna on top of the vehicles.

In table 3 the results after the decision fusion are shown. In sequence 3 the fusion improves results drasticly. The results for sequence 6 and 2 are better using a single sensor than those found after the fusion. This is due to the error on the image registration. For the calculation of Pd and Nft after fusion, all results are mapped in the coordinate space of the infrared sensor using image registration methods. Table 2 shows that the visual sensor in sequence 6 and 2 is almost solely responsible for all detections. Therefore, if the registration from the visual image to the infrared image is not accurate enough, a position declared as being a target by the visual sensor might just fall within the true target region while, when it is mapped into the coordinate space of the infrared image, it falls just outside and hence is counted as a false target.

7. Conclusions

In this paper an approach is presented for the detection of targets using multi-sensor image sequences. The developed algorithm consists of two parts for each sensor. The first part detects targets in single images and is based on texture measurements. For this part of the algorithm a semi-supervised approach is followed. For each sensor a test image is chosen and targets are indicated manually. Then linear discriminant analysis is used to find the combination of the different texture parameters that optimises the discrimination between target and background. The weights for these measurements are specific for each of the sensor types (Infrared/Visual). The same weights are then used for the actual target detection in other images of the same sensor type, even for images from different sequences. The results of the first part of the algorithm show the visual and infrared sensors to be complementary. The second part of the algorithm specifically focusses on the detection of moving targets. If a target is moving in a direction that does not coincide with the viewing direction of the sensors, this part of the algorithm gives good results. In any case, the number of false targets produced by this part of the algorithm is very low. Each part of the algorithm behaves as an expert indicating the possible presence of a target. The final decision of the algorithm is reached by fusing the results of the experts for the different sensors. Because the information to be fused is binary (either a target is detected or it is not), the decision fusion is based on a "k out of N" voting-rule. Currently, ways to assign a confidence measure to each expert's result are being investigated. This will allow more sophisticated decision fusion methods to be used.

REFERENCES

1. J. Knecht, L. Sévigny, C. Birkemark, R. Gabler, B. Hoeltzener-Douarin, , J. Haddon, W. Beck, R. Cusello, E. Oestevold, and L. Garn, *Detection of Target Formations in IR Image Sequences*, NATO/AC243/P3/RSG9, 1990.
2. T. Peli, L. Vincent, and V. Tom, "Morphology-based algorithms for target detection/segmentation," in *Proceedings on Architecture, Hardware, and Forward-Looking Infrared Issues in Automatic Target Recognition - SPIE - USA (Orlando)*, april 1993.

3. D. Borghys, C. Perneel, and M. Acheroy, "Long range automatic detection of small targets in sequences of noisy thermal infrared images," in *Proceedings on Signal and Data Processing of Small Targets - SPIE - USA (Orlando)*, 4-8 April 1994.
4. R. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE* **67**(5), pp. 786–804, 1979.
5. P. Verlinde, "Numerical evaluation of the efficiency of a camouflage system in the thermal infrared (in dutch)," Master's thesis, Katholieke Universiteit Leuven (KUL/ESAT), Leuven, 1989.
6. P. Verlinde and M. Proesmans, "Global approach towards the evaluation of thermal infrared countermeasures," in *Proceedings of Chracterization, Propagation and Simulation of Source and Backgrounds - SPIE - USA (Orlando)*, vol. 1486, April 1991.
7. *SPSS Professional Statistics*, SPSS Inc, Chicago, 1994.
8. C. Perneel, M. de Mathelin, and M. Acheroy, "Detection of important directions on thermal infrared images with application to target recognition," in *Proceedings of Forward Looking Infrared Image Processing - SPIE - USA (Orlando)*, 12-16 April 1993.
9. P. DeBeir, *Détection d'objectifs a longue distance dans une séquence d'images infrarouges*, Royal Military Academy, 30 Avenue de la Renaissance, B-1040 Brussels, 1992.
10. R. Y. Tsai and T. S. Huang, "Estimating three-dimensional motion parameters of a rigid planar patch," *IEEE-ASSP* **29**(6), pp. 1147–1152, 1981.
11. R. Morris, *Image Sequence Restoration using Gibbs Distributions*. PhD thesis, Dept. of Egineering and Trinity College, Cambridge University, 1995.
12. R. Morris and W. Fitzgerald, "Discontinuous motion and occlusion estimation - theory and application," in *Proceedings of the Iternational Conference for Young Computer Scientists, Beijing*, July 1995.
13. E. Waltz and J. Llinas, *Multisensor Data Fusion*, Artech House, Boston, 1990.
14. R. Antony, *Principles of Data Fusion Automation*, Artech House, Boston, 1995.