Long Range Automatic Detection of Small Targets in Sequences of Noisy Thermal Infrared Images

Dirk Borghys, Christiaan Perneel and Marc Acheroy

Electrical Engineering Department, Royal Military Academy, 30 Renaissance avenue, 1040 Brussels, Belgium.

ABSTRACT

In this paper, an approach to the automatic detection of vehicles at long range using sequences of thermal infrared images is presented. The vehicles in the sequences can be either moving or stationary. The sensor can also be mounted on a moving platform. The targetarea in the images is very small, typically less than 10 pixels on target. The proposed method consists of two independent parts. The first part seeks for possible targets in individual images and then merges the results for a subsequence of images. The decision for this part of the algorithm is based on temporal and spatial consistency of the targets through the considered image subsequence. The second part of the algorithm specifically focuses on finding moving objects in the scene. Clearly, as the sensor may itself be moving too, the effect of this motion on the images has to be eliminated first. This was done using a model based registration technique. The algorithm proposed in this paper was implemented and tested on a set of 7 image sequences obtained from different sensors under diverse operational circumstances. The images in these sequences are mostly highly cluttered and noisy. Results show the two parts of the algorithm to be quite complementary. For some sequences, the first part yields good results (high ratio of detection probability to false alarm rate) while, for other sequences, the second part gives the best results.

1. INTRODUCTION

Passive 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(see refs [1], [2], [3]). However, detection and tracking of small, low contrast vehicles in a highly cluttered environment still remains a very difficult task. This paper describes an approach to tackle this problem. The approach was implemented and then tested on a set of seven image sequences obtained from different sensors under diverse operational circumstances. The targetarea in the images is very small, typically less than 10 pixels on target. The vehicles can be either moving or stationary. In most of the sequences, the sensor was mounted on a moving platform. No prior knowledge about the sensor motion was available, making the detection of moving targets quite difficult. The images are mostly 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. An overview of the global approach is presented in section 2. The first part is split in two phases. The first phase which is explained in section three detects possible targets in single images. The second phase of part I (section 4) uses spatial and temporal consistency of target occurence through the sequence to reject false alarms. Section 5 describes the second part of the algorithm that specifically focusses on finding moving targets. The two last sections of the paper present the global results and conclusions.

2. OVERVIEW OF THE APPROACH

The proposed algorithm consists of two independent parts. The first part does not make any assumption on target motion. The only hypothesis is that the targets will contain parts that are warmer than their direct surroundings and thus present a minimum positive contrast over a limited surface. If knowledge about range and expected targetclass is available, the size and aspect ratio of the surface can be limited. In the first part of the

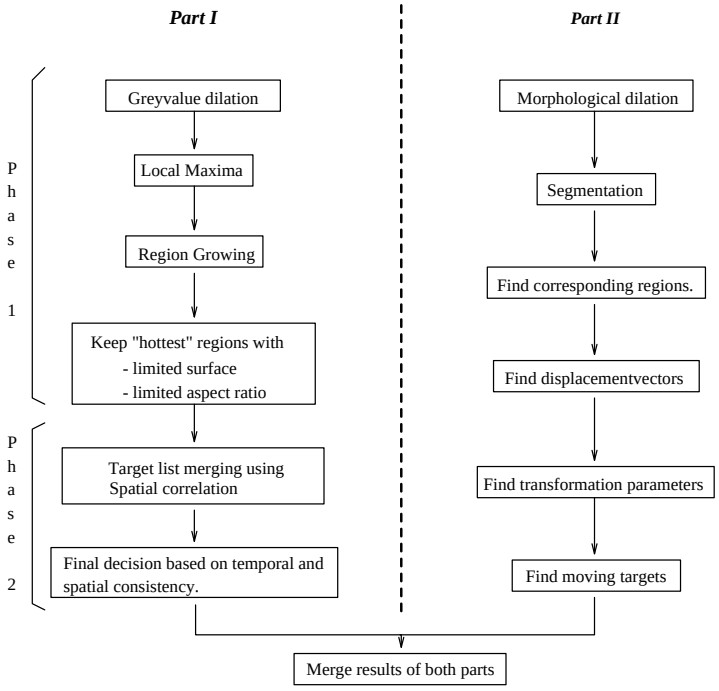


Figure 1: Overview of the approach

algorithm, *morphological operators* are used to select possible targets that match the hypotheses in the individual images. Then, the target lists are merged. Finally, a statistical test, based on the number of occurences of a possible target in the merged target list is used to reduce the false alarm rate.

The second part of the algorithm specifically tries to find the targets that are moving with respect to the background. Therefore, the effect of the moving sensor platform has to be eliminated. As the project is dealing with long range images, the terrain can be modeled by a *rigid planar surface*. In this case, the overall motion of the scene between two images can be expressed by eight motion parameters. These parameters are found by solving a system of linear equations obtained by using at least four displacement vectors. Once the motion of the background is modeled, the moving objects in the scene are detected. Temporal and spatial consistency of the target motion is used to eliminate false alarms.

The results of both parts of the algorithm are merged to give the final list of candidate targets. The proposed method was implemented and tested on a set of 7 image sequences called DIM01 to DIM07. In figure 1 a global overview of the approach is presented.

3. DETECTION OF POSSIBLE TARGETS USING SINGLE IMAGES (PART I, PHASE 1)

3.1. General Description of Phase 1

For the detection of targets in single images, a number of hypotheses are made. The first assumption is that the vehicles contain pixels that are warmer than their direct surroundings in the background and present a minimum contrast over a limited surface with these surrounding background pixels. Furthermore, it is assumed that the targets belong to the hottest points in the scene that present such a contrast. This phase of the algorithm is

implemented in two steps. In the first step morphological operators are used to find the local maxima. The second step takes into account the different hypotheses to find the possible targets among the local maxima.

3.2. Preprocessing

Before starting any processing on the images, for the first part of the algorithm, a greyvalue dilation is performed rescaling the greylevels of each image between 0 and 255.

3.3. Research of the Local Maxima

The procedure that determines the local maxima consists of morphological operators based on dilations with a centered horizontal element H of width 3 (cf. DeBeir [4] and Perneel [5]). The different steps are described and applied to a one-dimensional function in figure 2.

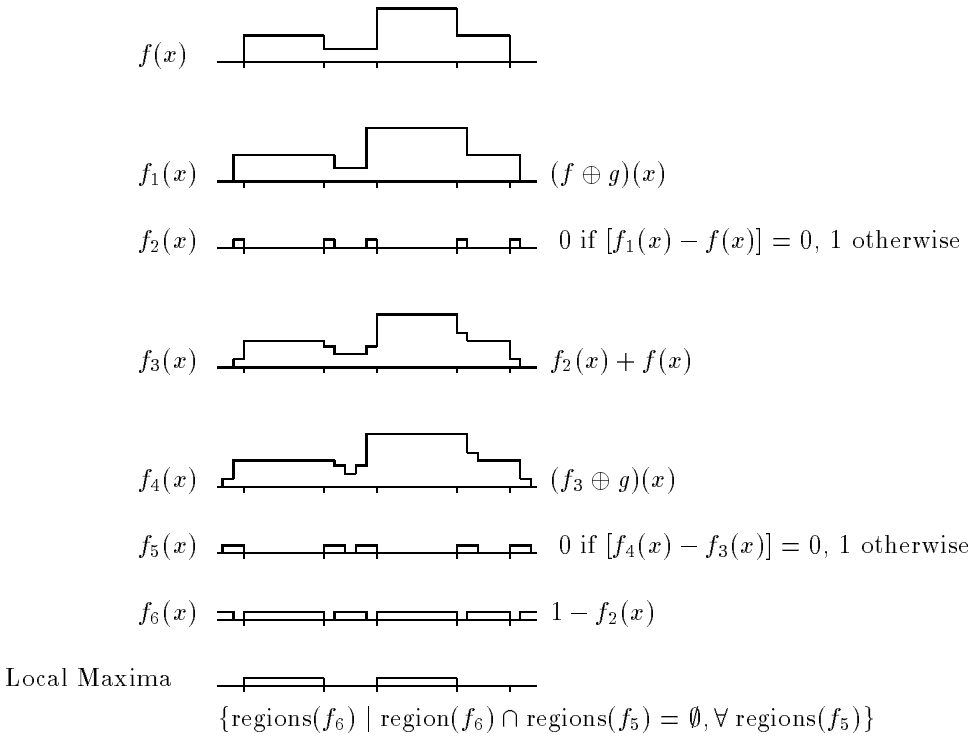


Figure 2: Detection of local maxima

3.4. Selection of the Possible Targets Amongst the Local Maxima

The selection of possible targets amongst the local maxima is implemeted as a region growing procedure that takes into account the different hypotheses (Fig 3). The parameters of the region growing are:

- The minimum contrast between target and background (H).

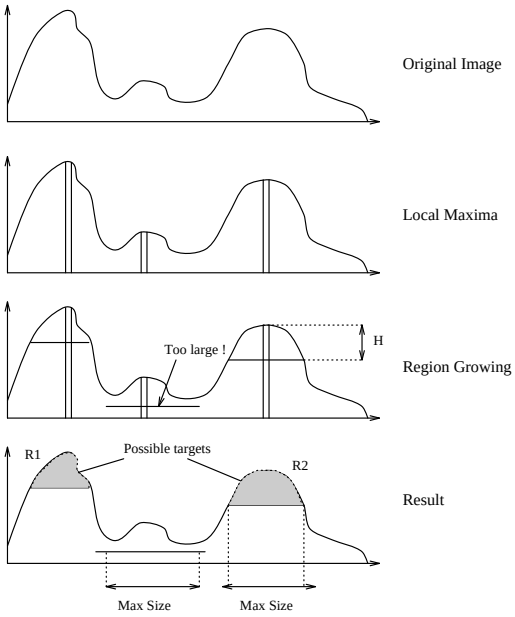


Figure 3: Region growing

- The minimum size of the target ($MinSize$)
- The maximum target size ($MaxSize$)
- The minimum and maximum aspect ratio for the regions ($MinAsp$ and $MaxAsp$)
- The maximum number of targets to be considered in each image ($MaxNrTargets$)

The region growing searches for the local maxima that can be grown into a region with a number of pixels between MinSize and MaxSize that have a maximum difference of H grey levels with the initial local maximum ($MaxGrey$). To take into account the hypothesis that the targets will belong to the hottest regions in the image, the search for valid regions starts at the local maxima with highest grey value and gradually descends until the number of valid regions is greater than $MaxNrTargets$.

4. REJECTION OF FALSE TARGETS USING MULTIPLE IMAGES(PART I, PHASE 2)

4.1. Merging of the Target Lists

The output of the first phase consists of files that contain the list of possible targets for each image I. The first step of Phase 2 is to merge these target lists. The merging of the target list I and $I+1$ is implemented by selecting a small window around each possible target in image I and finding the best match of this window in image $I+1$. The first step to find this best match consists of calculating the spatial correlation between the small window in image I and an equivalent window at the same coordinates in image $I+1$. Then, the position of the matching rectangle in the second image is spiraled out from its initial position and each time the spatial correlation improves by at least 10%, the matching vector is updated. Note that this matching algorithm only searches the best translation vector for the window. Therefore the matching is performed only between successive images where the possible effects of rotation of the sensor and scaling can be neglected. By chosing the matching window small

		DIM01	DIM02	DIM03	DIM04	DIM05	DIM06	DIM07
phase 1	Pd	95	51	25	92	82	0.5	18
	Nft	110	123	58	124	82	68	146
phase 2	Pd	79	38	13	89	59	0	14
	Nft	19	9	9	20	20	12	25

Table 1: Results of Part I of the algorithm for threshold = 80.

enough (e.g. 10×10 pixels) it is possible to track possible targets through the image sequence independently of the relative motion of the target with respect to the background. Once the optimal position of the matching window for a given possible target is found, this position is added to the target list of image $I + 1$. This merging step is repeated until the positions of all possible targets that were originally found in image I, are added to the target list of image $I + MaxDepth$. The parameter $MaxDepth$ thus represents the maximum length of the subsequences used in the second phase.

4.2. Selection of the Candidate Targets

After the merging of the target lists, the number of possible targets is enormous. However, clusters of target positions are found around the true targets (and around the details in the background that have the right dimensions and contrast with their surroundings). Thus the number of false alarms can be significantly reduced by performing a region growing around the possible targets in the merged list and then using a threshold to find the possible targets that are the most persistent along the considered part of the image sequence.

4.3. Results of Part I

The most critical parameter of the first part of the algorithm is the threshold. This is the maximum number of targets to consider in the first phase. To get an idea of the influence of this parameter, the detection probability and the number of false alarms after both Phase 1 and Phase 2 are plotted in figures 4a to 4d. The figures show that the number of false targets after Phase 1 increases almost linearly with the threshold. However, the probability of detection as a function of the threshold increases stepwise. At threshold value 80 the maximum is reached for Sequences 1,2,4 and 5. After Phase 2, the number of false alarms still increases almost linearly but the actual number is greatly reduced. The detection probability again shows a maximum around threshold 80.

To get an algorithm that is independent of the images, all parameters in Phase 1 and 2 were fixed for all sequences. The chosen values offer a compromise between the probability of detection and the number of false targets for all the image sequences. If supplementary information, such as range, were available, parameters could be set more finely and results would improve significantly.

The results of the calcultations with threshold 80 are shown in Table 1. For each sequence the probability of detection and the average number of false targets per frame is given.

In four sequences a detection probability between 50 and 95 % is obtained after Phase 1. In DIM06 none of the targets is detected due to the poor image conditions (high clutter and noise combined with small, fast moving targets). In DIM07 the targets are only detected in the first seven images. At image 7 an explosion occurs which causes all declared targets in subsequent images to be located at the edges of the smoke cloud. The second phase offers a significant reduction of the false alarms rate (typically 70%). In most of the sequences (DIM01, 02, 04, 05 and 07), the probability of detection is reduced by 20 to 30% in the second phase. However in DIM03 almost 50% of the targets are lost in the second phase. This is due to the fact that the targets in DIM03 are only sporadicly detected by Phase 1.

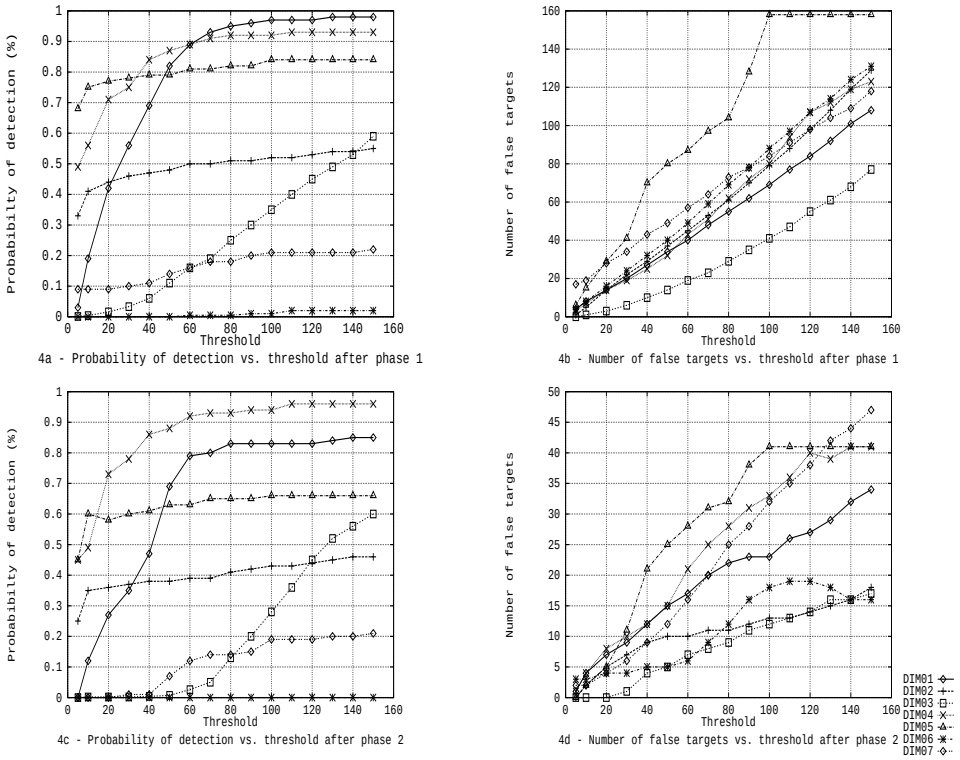


Figure 4: Results of part I

5. IMAGE REGISTRATION AND SEARCH FOR MOVING TARGETS (Part II)

5.1. Introduction

The purpose of this algorithm is to find targets that are moving with respect to the background. Evidently, the images have to be registered first. This is done using a model for the global transformation of the background due to the sensor motion. The moving targets will be objects that do not correspond to the model.

5.2. General Description of the Method

After compairing some image registration techniques (see refs [6] [7] [8] [9] [10]), the method that seemed the most appropriate for this application is based on an article by Tsai and Huang [11]. In this article the three-dimensional motion parameters of a rigid planar patch are estimated using displacement vectors between two time-sequential image frames (see also Fig 5).

Let x, y, z be the coordinates in space of a point P at time $t1$
Let x', y', z' be the coordinates in space of a point P at time $t2$
Let X, Y be the image coordinates of the point P at time $t1$
Let X', Y' be the image coordinates of the point P at time $t2$

Then, if z is chosen along the optical axis of the sensor:

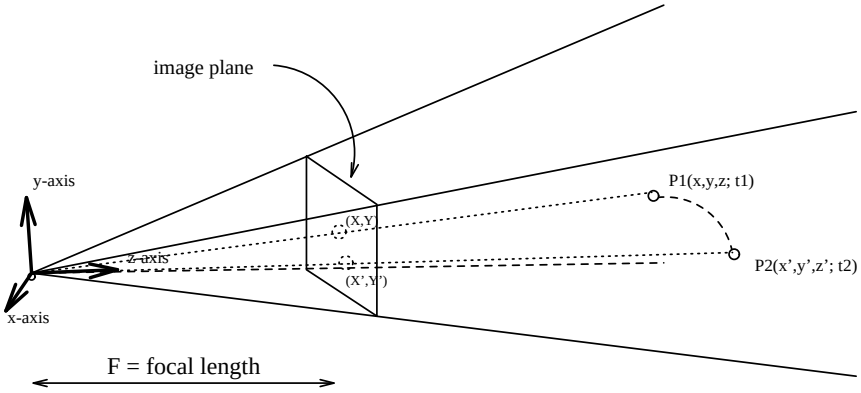


Figure 5: Transformation of the scene

$$\begin{array}{ll} X = Fx/z & X' = Fx'/z' \\ Y = Fy/z & Y' = Fy'/z' \end{array} \tag{1}$$

In which F is the focal length of the sensor system.

If, between time $t1$ and $t2$ the object has undergone translation and rotation, the general transformation matrix is given by:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & -r_3 & r_2 \\ r_3 & 1 & -r_1 \\ -r_2 & r_1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \tag{2}$$

The restriction in the article is that only points in a planar patch are considered in determining the transformation parameters. As this project is dealing with long range images, this is not an important limitation. Considering only points in a plane means that:

$$ax + by + cz = 1 \tag{3}$$

then :

$$z = \frac{F}{aX + bY + F} \tag{4}$$

and :

$$\begin{aligned} X' &= \frac{a_1 X + a_2 Y + a_3}{a_7 X + a_8 Y + 1} \\ Y' &= \frac{a_4 X + a_5 Y + a_6}{a_7 X + a_8 Y + 1} \end{aligned} \tag{5}$$

with :

$$\begin{array}{ll} a_1 = \frac{1+a\Delta x}{1+c\Delta z} & a_2 = \frac{-r_3+b\Delta x}{1+c\Delta z} \\ a_3 = \frac{(r_2+c\Delta x)F}{1+c\Delta z)} & a_4 = \frac{r_3+a\Delta y}{1+c\Delta z} \\ a_5 = \frac{1+b\Delta y}{1+c\Delta z} & a_6 = \frac{-(r_3+c\Delta y)F}{1+c\Delta z} \\ a_7 = \frac{-r_2+a\Delta z}{(1+c\Delta z)F} & a_8 = \frac{r_1+b\Delta z}{(1+c\Delta z)F} \end{array} \tag{6}$$

A mapping between the two images will be completely defined by the eight parameters $a_1 ... a_8$ of equation 5. This means that one needs to find at least 4 displacement vectors to be able to find the global transformation

between two images. If only 4 corresponding points are used no three of them may be colinear[12]. Having found these displacement vectors, the transformation is determined by solving the system of linear equations in equation 5. Moving targets will be among the regions in the image that do not conform to the tranformation model.

5.3. Implementation

Figure 6 shows a block diagram of the approach used to find the transformation parameters. In the following the different steps are discussed in detail.

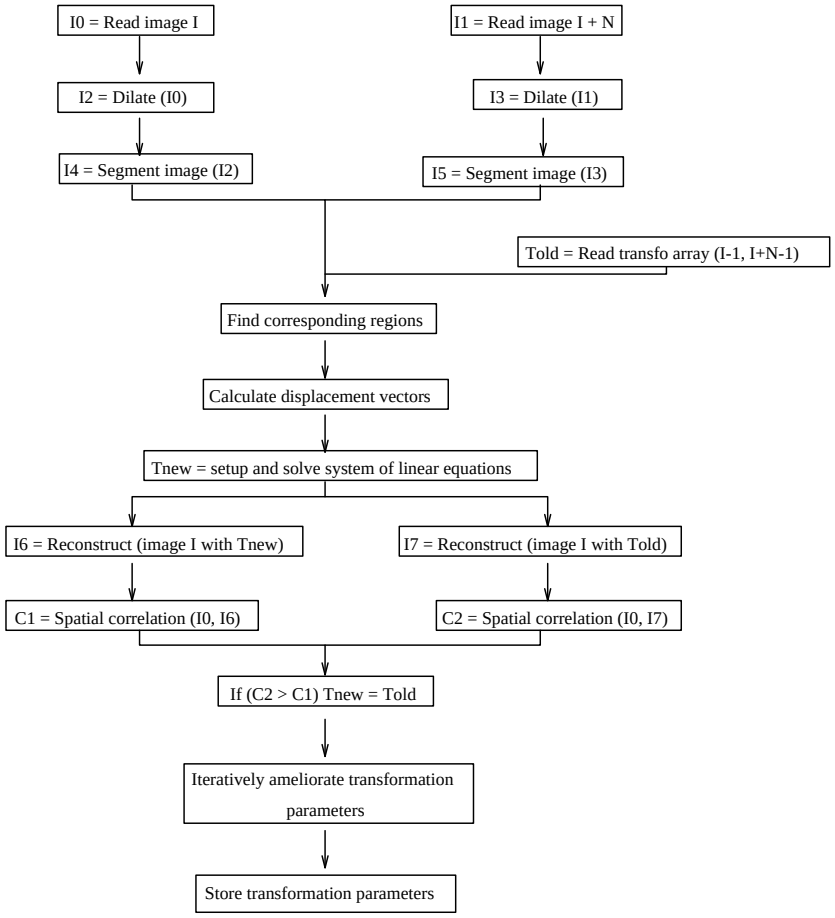


Figure 6: Method used to find the transformation of the scene

5.3.1. Preprocessing and Segmentation The segmentation used to find the displacementvectors is based on the fact that it is easy to find and label large, hot (bright) regions in a scene. To get larger, homogeneous regions with high grey levels, the images are preprocessed by applying a morphological dilation using a 5×5 kernel. The segmentation is then performed by applying the largest threshold that results in a given number of regions (N_{wanted} = 15) with a number of pixels between N_{min} (=20) and N_{max} (=10000).

5.3.2. Find the Corresponding Regions Using the threshold to segment the images, the regions with highest apparent temperature are found for each image. It is therefore expected that in each image more or less the same regions will be selected. However, as the grey value in the images may change, the dimension and shape of the

regions are not constant. The only information that can therefore be used to find the corresponding regions in two images is the relative position of those regions.
For each region in both images the distance and angles to each other region in the same image is calculated and stored in a vector. If region R_k of image I corresponds to region R_l of image $I+N$, it is expected that for each distance and angle in the vector of region R_k, a similar distance and angle can be found in the vectors of region R_l of image $I+N$. For each distance and angle in the vector of R_k the best matching distance and angle for the considered region in image $I+N$ is found and the sum of the absolute differences for both the distances (sumdist) and the angles (sumangle) is calculated. The region in image $I+N$ with the smallest sumdist and sumangle is the region that best corresponds to region R_k of image I. To determine the third criterion used to find the corresponding regions between images I and $I+N$, the transformation matrix that has already been found between image $(I-1)$ and $(I+N-1)$ is used. This transformation is applied to the center of mass of region R_k and the result is expected to lie close to the center of mass of the corresponding region. The third parameter is thus the distance between each region and this predicted location. The regions in image $I+N$ are now ordered in terms of increasing value of the three parameters. The selected region is the one with the smallest sum of ranks. The selected region in image $I+N$ is validated using the product of the three parameters. This product should be small for a valid match (typically < 100).

5.3.3. Set Up and Solve the System of Linear Equations The centers of mass of the validly matched regions are now used as samples of the displacement field between both images and the system of linear equations is set up. If at least four displacement vectors are available, Gaussian elimination is used to solve the system. Solving the system of equations provides a first approximation of the transformation parameters $a_1 ... a_8$. If less than four valid displacement vectors were found, the parameters found for the previous couple of images is used in the next step.

5.3.4. Refine the Transformation Parameters Because the solving of the system was only based on the displacement vectors of the mass centers of a few corresponding regions, the transformation parameters found so far should be looked upon as a first approximation. The eight transformation parameters are now iteratively refined by optimizing the spatial correlation between image I and its reconstruction from image $I+N$ found by using equation 5.

5.3.5. Finding the Moving Targets The method used to find the moving targets is shown in Fig 7. The transformation parameters are used to reconstruct image I from image $I+N$. If the transformation was determined accurately, the differences between the original and the reconstructed image consist of noise and of points that do not match the model of a moving rigid planar surface, i.e. points belonging to moving objects or to objects that are not in the plane formed by the rest of the scene.

At a position of a moving target it is expected to find a region of minimum difference (the threshold) in grey value between the original image and the reconstruction (R1) and closeby, a region that has a similar grey value and size but has the inverse difference in grey value (R2). Only couples of such nearby regions are considered as being possible moving targets. Because vehicles contain points that are warmer than their surroundings, for each couple of regions, the one with the highest grey value (P2) in the reconstructed image will yield the target position in image $I+N$ (P4) (after applying the transformation) while the other region gives the position in image I (P1). If the couple of regions corresponds to a moving target, it should be possible to find the complete path. A possible moving target is discarded if the final point (P4) of the path is not close to the predicted location (P3) in image $I+N$.

To limit the number of false alarms, a statistical test is used to accept only the target positions that are present in a sufficient number of target paths.

5.4. Results of Part II

The transformation is accurately found in all sequences except DIM05 where not enough regions can be identified in the background and DIM07 where the matching is perturbed by the explosion cloud. Figures 8a and 8b show the probability of detection and the number of false alarms for this part of the algorithm in function of the threshold.

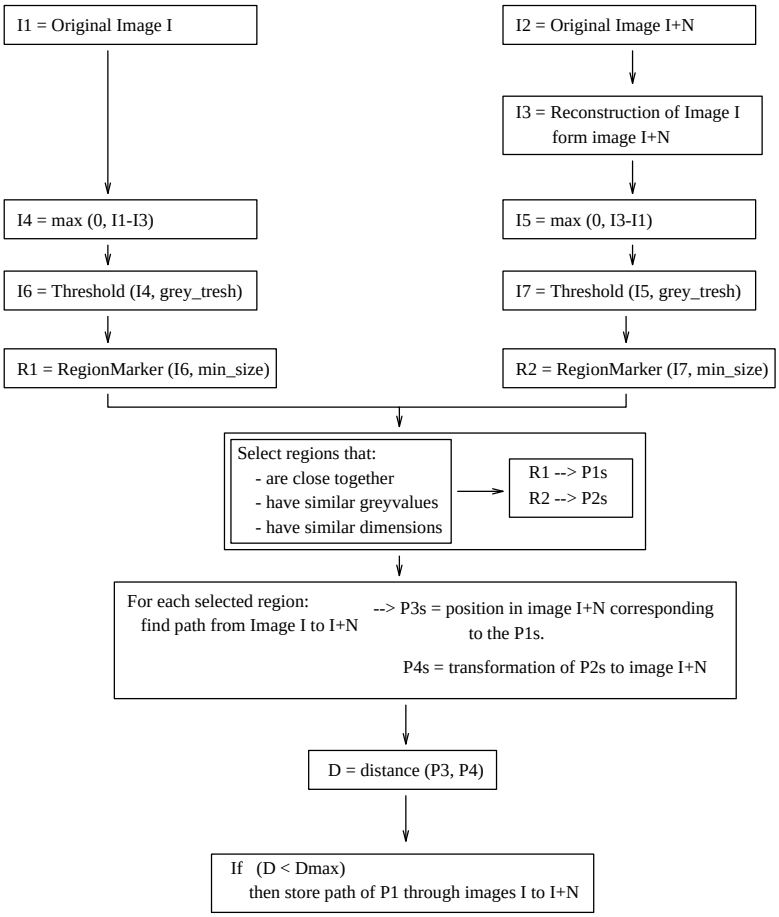


Figure 7: Method used to find the moving targets

Note that the second part of the algorithm gives good results for DIM06 where the first part fails to detect any targets. The probability of detection falls drasticly between thresholdvalue 25 and 30 for DIM04 and DIM06.

6. Global Results

In figures 9a-d the results of the global algorithms in function of the threshold of part I are presented. For these plots, four values of the first treshold were used (40,60,80,100). In each plot the detection probability is given versus the number of false targets. In each of the plots, the threshold for the second part of the algorithm is fixed. The treshold of the second part is 35, 30, 25 and 20 in figures 9a, 9b, 9c and 9d respectively. In some sequences (01,04 and 05) the first part of the algorithm gives very good results. In those sequences the second part of the algorithm does not add a lot of false targets and does not increase the probability of detection significantly.

In two sequences (DIM03 and DIM06) however, the second part largely contributes to the number of detected targets. In DIM06 the second part is even solely responsable for all the detected targets. In DIM07 an explosion occurs in Image 7; in all subsequent images a bright smoke cloud is visible which contains so many false alarms for the first part that no targets are detected after the explosions. The explosion cloud also prevents the second part of the algorithm to find the right interframe displacement vectors so that it is impossible to model the global scene transformation. Even if it would be possible to find the correct scene transformation, all moving targets found by part two of the algorithm will probably be situated in the expanding smoke cloud. For DIM01 to DIM06 the two parts of the algorithm seem to be complementary. For each image sequence it is possible to find a combination of both thresholds that yields an optimal detection probability/number of false targets rate. Table 2 shows the results

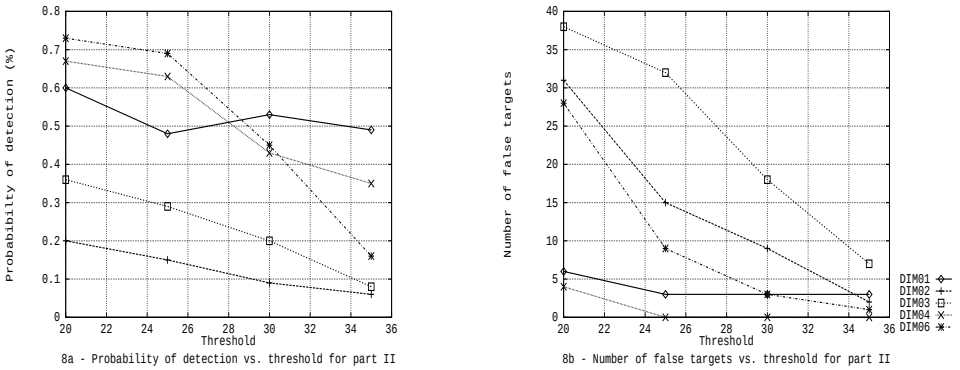


Figure 8: Results of part II

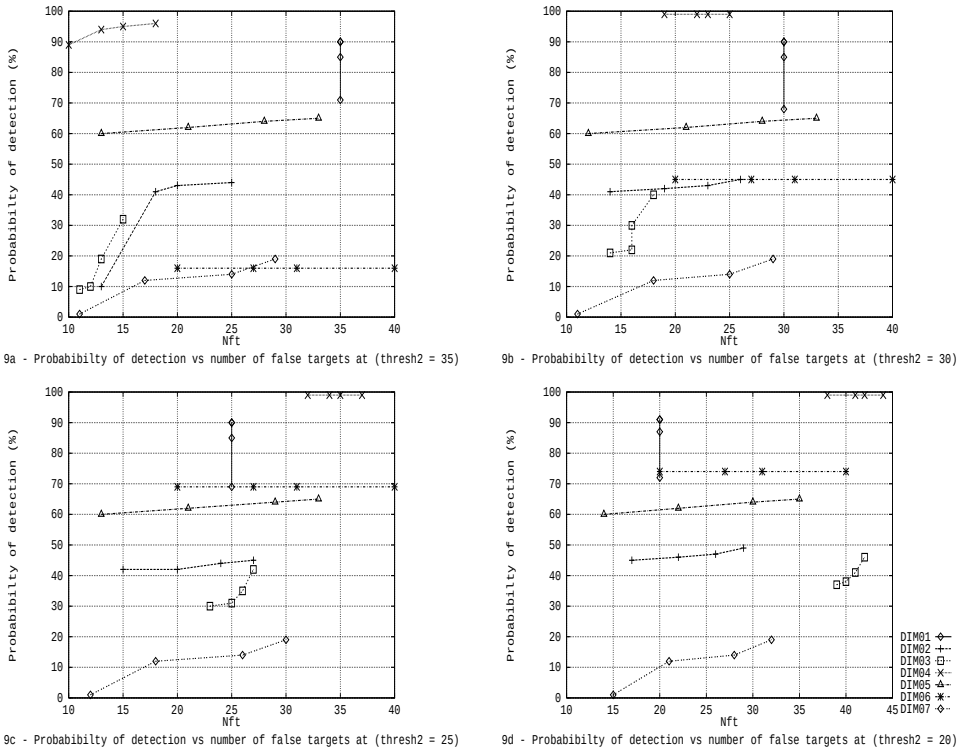
Sequence	Threshold 1	Threshold 2	P_d	N_{FT}
DIM01	80	35	0.90	20
DIM02	100	35	0.44	15
DIM03	140	35	0.57	25
DIM04	40	30	0.99	12
DIM05	80	-	0.59	20
DIM06	0	25	0.69	9
DIM07	100	-	0.19	32

Table 2: Results with a sequence dependent choice of the thresholds

for a sequence dependent choice of thresholds. It may be possible to use a priori knowledge about the terrain under observation and the properties of the sensor to set the optimal thresholds (and other parameters) prior to or during the mission.

7. Conclusions

In this paper an approach to the automatic detection of vehicles at a long range using sequences of thermal infrared images is presented. The algorithm is divided into two parts. In the first part no assumptions are made about target motion. The detection is based on the temporarily consistent presence of hotspots. The second part explicitely searches for possible targets that are moving with respect to the background. This second part of the algorithm contains an approach to eliminate the apparent interframe motion of the background which is caused by a moving sensor. Analysing the global results, it appears that the two parts of the algorithm are complementary. It should be noted that the results could be dramatically improved if more knowledge about the operating conditions at the time of scene observation were available. Knowledge of the motion, position and orientation of the sensor and of the type of background would indeed enable one to tune the different parameters (e.g. possible target size) of the algorithm.



9a - Probabibilty of detection vs number of false targets at (thresh2 = 35)

9b - Probabibilty of detection vs number of false targets at (thresh2 = 30)

9c - Probabibilty of detection vs number of false targets at (thresh2 = 25)

9d - Probabibilty of detection vs number of false targets at (thresh2 = 20)

Figure 9: Final Results

8. REFERENCES

[1] J. Knecht, L. Sévigny, C. Birkemark, R. Gabler, B. Hoeltzener-Douarin, , J. Haddon, W. Beck, R. Cusello, E. Oestevold, and L. Garn, "Discrimination and classification of operating military targets in natural scenes from thermal imagery," *Computer Vision, Graphics and Image Processing*, vol. 11, pp. 185–191, 1979.

[2] 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.

[3] 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.

[4] P. DeBeir, *Détection d'objectifs a longue distance dans une séquence d'images infrarouges.* 30 Avenue de la Renaissance, B-1040 Brussels: Royal Military Academy, 1992.

[5] 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.

[6] L. G. Brown, "A survey of image registration techniques," *ACM Computer Surveys*, vol. 24, no. 4, pp. 325–376, 1992.

[7] R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces," *IEEE-PAMI*, vol. 6, no. 1, pp. 13–27, 1984.

[8] W. Burger and B. Bhanu, "Estimating 3-d egomotion from perspective image sequences," *IEEE-PAMI*, vol. 12, no. 11, pp. 1040–1058, 1990.

[9] H. H. Nagel, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE-PAMI*, vol. 8, no. 5, pp. 565–593, 1986.

[10] H. J. Lee and H. C. Deng, "Three-frame corner matching and moving object extraction in a sequence of images," *Computer Vision, Graphics and Image Processing*, vol. 52, pp. 210–238, 1990.

[11] R. Y. Tsai and T. S. Huang, "Estimating three-dimensional motion parameters of a rigid planar patch," *IEEE-ASSP*, vol. 29, no. 6, pp. 1147–1152, 1981.

[12] J. Weng, T. S. Huang, and N. Ahuja, *Motion and Structure from Image Sequences.* New-York: Springer Verlag, 1993.