TARGET DETECTION FROM COREGISTERED VISUAL-THERMAL-RANGE IMAGES

Jorge E. Pérez-Jácome and Vijay K. Madisetti

Center for Signal and Image Processing (CSIP), Georgia Institute of Technology – ECE, Atlanta, GA 30332, USA
e-mail: jorge@eedsp.gatech.edu, vkm@eedsp.gatech.edu

ABSTRACT

A method to automatically detect targets from sets of pixel-registered visual, thermal, and range images is outlined. It uses operations specifically designed to work on the different kinds of images to explote the information given by each of them. Five features are used to distinguish the targets from the clutter: texture, brightness, temperature, surface planarity, and height. The results from individual detectors are then combined to improve the detection rate while reducing the number of false alarms. A morphological operation called "erosion of strength n" is also introduced and utilized as a powerful tool for removal of spurious information. The excellent results obtained for detection support the suitability of this approach for other ATR (Automatic Target Recognition) problems.

1. INTRODUCTION

The reason for using several images to detect and recognize targets is that we can take advantage of the different kinds of information represented by them in order to raise the detection and/or recognition rates and to reduce the false alarm rate. A very good, yet brief, presentation of the different sensors used on ATR is given by Bhanu and Jones [1]. Work on target detection from multsensor images has mainly dealt with only two input images [5, 4]. Our approach here is to define three basically different kinds of images according to what they represent, without concern for the specific method and/or sensor used to produce them:

Visual Images: Those that represent the intensity of the light emitted or reflected by bodies, within the visible band of the spectrum. A regular photograph is the typical example of this kind.

Thermal Images: Those whose pixel values represent a measure of the temperature at a specific location. Actually, they represent the intensity of light emitted or reflected by bodies, but inside a certain infrared region of the spectrum. Under certain conditions, the intensity obtained from an infrared [8–12μm] sensor is precisely related to the exact temperature by a simple line equation [2].

Range Images: Those whose pixel values represent a measure of the distance from the objects to the sensor. On top-view aerial images, these images can represent elevation of terrain or objects.

This work was supported in part by the U.S. Army Research Laboratory under Cooperative Agreement DAAL01-96-2001.

Although the selection of a set of images to perform ATR is non-unique and is certainly not a simple task, we believe that the three kinds of images mentioned above form a sufficiently orthogonal basis on which to perform target detection, based on the inherent differences among them.

The methods to produce the images can be very diverse: The sensors can be active or passive, they may use a given specific band or another. Even for a single kind of image, several methods could have been employed for its generation, but the resultant images are of the same nature, and so, can be operated on by algorithms defined for the specific kind of image. In the following section, we describe a method to perform target detection from sets of three pixel-registered images (visual-thermal-range) for a given scene.

2. DETECTION ALGORITHM

The general scheme for the detection of targets from visual-thermal-range image sets is presented in figure 1. The system is designed to operate on top-view images with pixels represented by bytes (0 to 255). On the visual images, higher values represent brighter points. On the thermal images, higher values represent warmer points. The range images follow a format in which one-level increments correspond to changes of 10 cm in elevation. The images must be pixel-registered (any point in a scene represented by the same coordinates in the three images). The resolution for the images is 25 cm per pixel, and the targets have rectangular to elliptical shapes, with of area 150 to 2000 pixels.

Targets are characterized by a set of features that discriminate them from the clutter: Visual texture and brightness, temperature, surface smoothness, and height. Based on these features, we develop five different target extractors specifically designed to look for regions with these characteristics to perform their operations. Upon cleaning of spurious information, the detection information from each of these extractors is combined to obtain the overall detection. With this scheme, targets that are missed from a given image can be can be detected by the general system, whereas false alarms produced by individual detectors are not present in the integrated detector. The implementation of each of the blocks of figure 1 is described next.

2.1. Bright/Dark point Extractor

The bright/dark-point extractor is used on both visual images and thermal images. It extracts points that are either

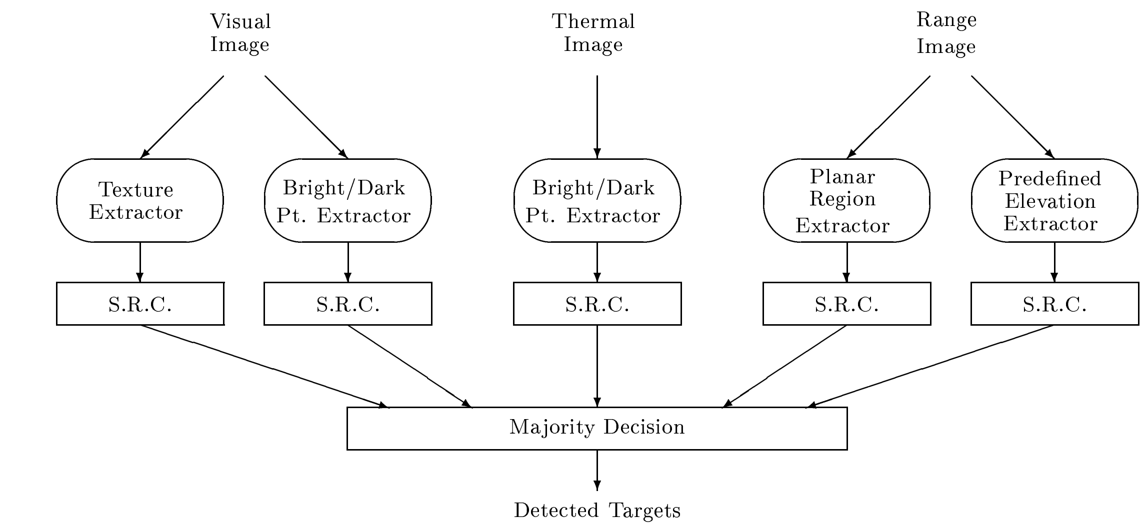

Figure 1: The detection algorithm

darker or brighter than their surroundings in visual images, and points that are either warmer or colder than their surroundings in thermal images. Our method makes use of a *rectangular annular window* to discriminate between targets and clutter. This window, one pixel wide, is placed around a given test pixel (i, j) (presumably a target) as shown in figure 2, and estimates the mean $\mu_{i,j}$ and standard deviation $\sigma_{i,j}$ of its pixels. Then, to determine the possibility that the test pixel is part of a target, it checks whether its value $x_{i,j}$ differs from $\mu_{i,j}$ by more than $1.5\sigma_{i,j}$. That is,
if $(x_{i,j} - \mu_{i,j}) > 1.5\sigma_{i,j}$ or $(x_{i,j} - \mu_{i,j}) < -1.5\sigma_{i,j}$,
then assign point (i, j) as a possible target.

2.2. Texture Extractor

The texture extractor operates on visual images. It measures the degree of similarity between adjacent pixels, for both the point under study (i, j) (presumably a target) and the pixels on an annular window around it (presumably clutter), and then compares them to see if they differ by more than a specified amount. As in the bright/dark-point extractor, we calculate a mean and a standard deviation for the annular window, but of the absolute difference between adjacent pixels, rather than of their intensity. Also, we calculate $x^{\Delta}_{i,j}$, the average difference in value between point (i, j) and its four adjacent points. So the test to determine a target point becomes:
if $(x^{\Delta}_{i,j} - \mu^{\Delta}_{i,j}) > 1.5\sigma^{\Delta}_{i,j}$ or $(x^{\Delta}_{i,j} - \mu^{\Delta}_{i,j}) < -1.5\sigma^{\Delta}_{i,j}$,
then assign point (i, j) as a possible target.

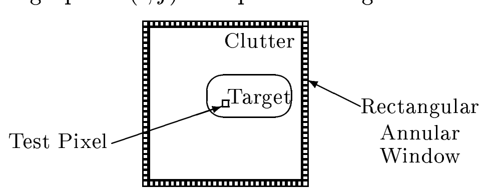

Figure 2: Annular Window for Target Detection

2.3. Planar Region Extractor

Since targets are well modeled by a collection of planar regions, the use of the degree of planarity to determine possible targets has been proposed [4]. A target usually has smooth (planar for small regions) surfaces, compared to most forms of clutter (grass, trees, ground). The planar region extractor examines 3×3 pixel regions from the range images, and obtains an error e with respect to the equation of a plane, $z = ax + by + \rho_0$. The error is estimated using the equation

$$e = ||\mathbf{X}\mathbf{w} - \mathbf{z}||^2, \text{ where}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{x_1}^t \\ \mathbf{x_2}^t \\ \vdots \\ \mathbf{x_9}^t \end{bmatrix}, \qquad \mathbf{x_i} = \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}, \qquad \text{and} \quad \mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_9 \end{bmatrix}$$

The elements of of $\mathbf{x_i}$ include the x and y coordinates for each of the 9 points under study. Similarly, $\mathbf{z}$ represent the heights for those pixels derived from the range values. $\mathbf{w}$ is a 3-element vector representing the best fit (with minimum square error) of the 9 points to a plane;

$$\mathbf{w} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{z}$$

The extractor uses a threshold $e_{TH} = 0.6$; a pixel is defined as a target if $e < e_{TH}$.

2.4. Predefined Elevation Extractor

If we have a basic knowledge of the kind of targets to search for (in our case, tanks), we can easily check if a point under study has an elevation suggesting a possible target. For this, we calculate $\mu_{i,j}$, the average elevation of a surface in an annular window around a point (i, j), and then compare it with $x_{i,j}$, the elevation of the point (i, j):
if 80cm $< (x_{i,j} - \mu_{i,j}) <$ 250cm ,
then assign point (i, j) as a possible target.

Table 1: Detection results; (√) detected, (×) miss.

		Scene 1			Scene 2				Scene 3				SRC parameters	
Image	Extractor	t_1	t_2	FA's	t_1	t_2	t_3	FA's	t_1	t_2	t_3	FA's	n_1	n_2
Visual	Texture	√	×	9	√	√	√	28	√	×	×	16	3	5
Visual	Bright/Dark	√	×	9	√	×	×	13	√	×	×	8	5	8
Thermal	Bright/Dark	√	√	13	√	√	√	13	√	√	√	21	5	8
Range	Pr. Elevation	√	√	6	√	√	√	10	√	√	×	17	4	7
Range	Planarity	√	√	24	√	×	√	21	√	√	×	12	4	5
Overall Detection Results		√	√	4	√	√	√	5	√	√	×	7	- - - - -	- - - - -

2.5. Spurious Region Cleaner (S.R.C)

The output of each of the extractors defined above is a *binary* image (1: target, 0: no target), downsampled by 4:1 (after or during the extraction operation) to reduce complexity while maintaining most of the detection information. A problem with those images is that they present many isolated single-pixel blobs, or some blobs that definitely do not have the shape of a target. Also, there are some blubs that can be recognized by eye as targets, but which have many "holes" (pixels with value 0) in them. To remove this "noise," we use a series of morphological operations that are specifically designed for this purpose. We propose an erosion operator *eros-n(im,n)*, "erosion of strength n" which works as follows: a 3×3 template is passed over the binary image im. Around each pixel (i, j), the number of 1's is counted. If it is larger than n, the output for the pixel (i, j) is 1, otherwise, it is 0. This operator can be used in series, with different values of n, and gives excellent results. This powerful but simple operator can be stated with MATLAB code as follows:

```
b=[1 1 1; 1 1 1; 1 1 1];
imOUT=conv2(imIN, b, 'same');
imOUT= (imOUT >= n );
```

When $n = 9$, it degenerates into the classical 3×3 erosion operator. By using $n < 9$, we keep points on the input image that are important, but that would be eliminated with other erosion methods. Note also that the operator is independent of shape. An analogous "dilation of strength n" can also be defined. The spurious region cleaning (SRC) operator is defined as follows:

$$im_{out} = dilate(eros\text{-}n(eros\text{-}n(im_{in}, n_1), n_2)$$

The application of two erosion operators in series results in great performance on eliminating spurious target pixels, for different densities and target-to-clutter contrasts. The parameters n_1 and n_2 are toned to specifically work on the different images. The *dilate* operator is used to join together points that are likely to belong to the same target. The output gives small regions inside the likely targets, usually with very few false alarms, for the different binary images.

2.6. Majority Decision

The final function of the detector is to combine the results of the individual detectors to produce the final output. The method we use checks the five detectors, and for each pixel it assigns a 1 if there are three or more 1s as inputs, and assigns a 0 otherwise. If a cluster of 1s overlaps a target, the target is declared *detected*, otherwise it is declared a *miss*. A cluster not overlapping a target is declared a *false alarm* (FA).

3. TEST DATA AND RESULTS

We analyzed three different sets of images (each consisting of a visual, a thermal, and a range image), representing three scenes. The first scene has two tanks, on a dry area, without vegetation. The second scene has three tanks, including one partially occluded by vegetation. The third scene has also three tanks including one partially occluded by vegetation, and it has several pieces of cultural clutter, such as small buildings, bridges, etc. The last two scenes have bodies of water as well. The images, 512×512 pixels, are artificial, but were synthesized with information from real visual images. [1]

The generation process was as follows: Visual backgrounds were taken from selected aerial photographs. These images were clipped and scaled to match our objectives. Then, we embedded visual images of tanks on the images, with the use of interactive programs. The location and orientation of the targets were chosen to resemble a real scene as closely as possible. Then thermal images were first generated with the use of interactive tools to define temperature values for every part of the images, and then post-processed with the use of filtering, interpolation, and the addition of spatially correlated random data. Finally, the range images were synthesized in a similar way, incorporating not only the elevation data, but the random height variability of the different surfaces that composed the scene, as would result from subpixel information.

As seen in table 1, some targets could not be detected from individual images, but most were correctly detected by our integrated system. Of the total of eight targets, seven were correctly detected, and only 16 false alarms were produced, almost half of them from cultural clutter in the third scene. The number of false alarms per scene was smaller than the number of false alarms obtained for any individual detector. The only *miss* corresponds to a partially occluded target. A series of images with the complete detection results for scene 2 is shown in figure 3; t_1, t_2 and t_3 are in the left, bottom left, and bottom right respectevely, t_2 is partially occluded by foliage. Note how the different values (n_1, n_2) used for the double SRC operator make it possible to remove spurious information from very different concentrations and distributions of possible target pixels. We plan

[1] Thermal Images have been modified to avoid the over-ease of detection presented in [6].

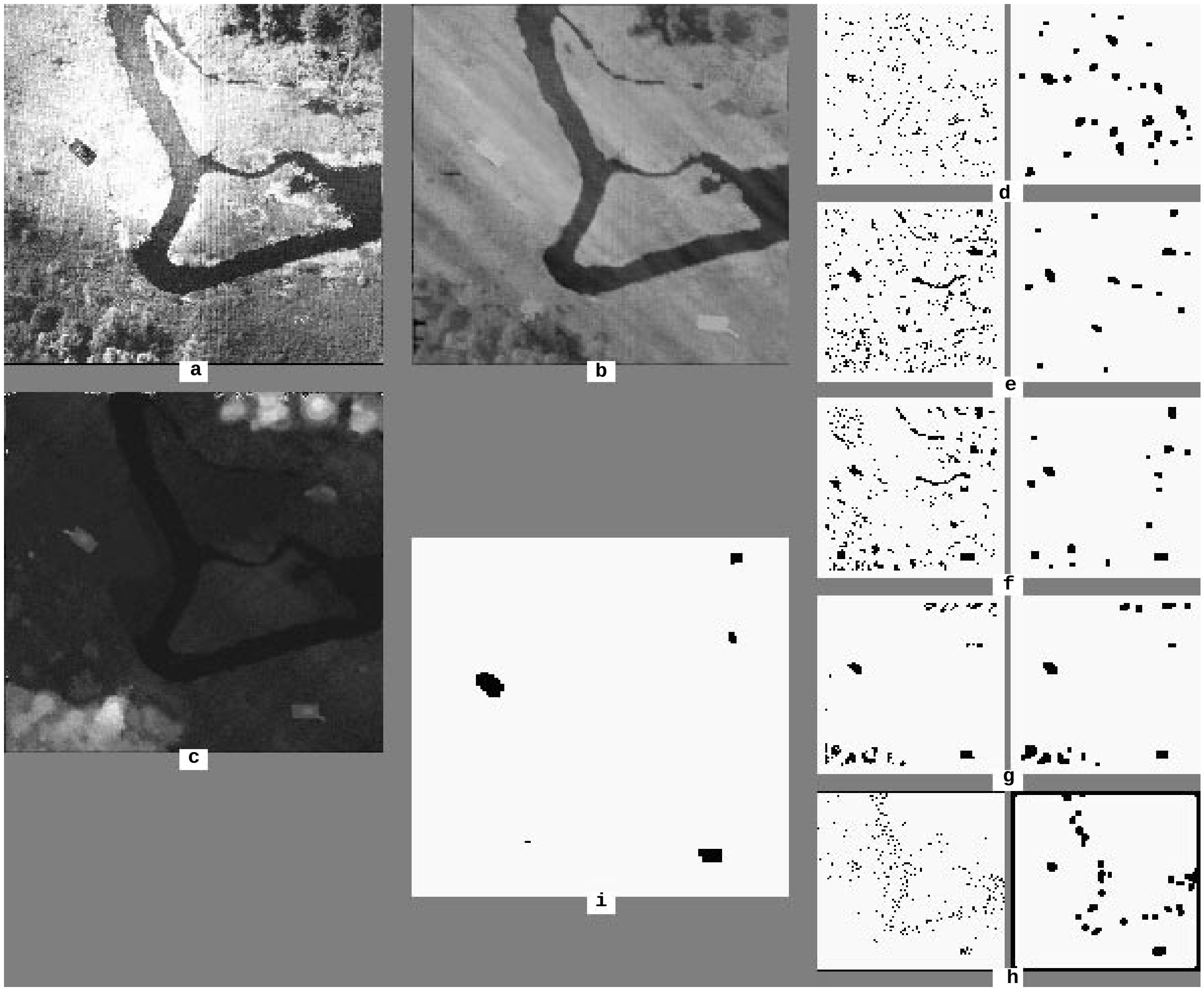

Figure 3: Detection process for scene 2. (a-c) Original visual, thermal, and range images. (d-h) show the detection before SRC (left) and after SRC (right), as they result from: (d) texture extractor on visual image, (e) bright/dark extractor on visual image, (f) bright/dark extractor on thermal image, (g) predefined elevation extractor on range image, (h) planar region extractor on range image. (i) overall detection results.

to apply our system to additional real scenes when data becomes available, to see its actual performance and to adjust the parameters correspondingly. We think that the process of classification can be greatly improved by the use of this scheme. Once the targets are detected, direct template matching for the different images can be easily applied.

4. REFERENCES

[1] B. Bhanu and T. L. Jones, "Image Understanding Research For Automatic Target Recognition," IEEE Aerospace and Electronic Systems Magazine, Vol. 8, No. 10, pp. 15–23, October, 1993.

[2] Nandhakumar and Aggarwal, "Integrated Analysis of Thermal and Visual Images for Scene Interpretation," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 10, No. 4, pp. 469–481, July, 1988.

[3] J. Nahm, "Joint ATR-Compression for FLIR and SAR Images," pr. doctoral dissertation, Georgia Tech 1996.

[4] M. C. Roggemann et al., "An Approach to Multiple Sensor Target Detection", SPIE Sensor Fusion II, Vol. 1100, pp. 42–52, 1989.

[5] R. L. Delanoy et al., "Machine Intelligent Automatic Recognition of Critical Mobile Targets in Laser Radar Imagery", The Lincoln Laboratory Journal, Vol. 6, No. 1, pp. 161-186, 1993.

[6] J. E. Pérez-Jácome and V. K. Madisetti, "Automatic Target Detection from Pixel-Registered Visual-Thermal-Range Images", submitted to the 1997 ARL Sensors and Electron Devices Symposium.