

# Aerial Moving Target Detection Based on Motion Vector Field Analysis

Carlos R. del-Blanco<sup>1</sup>, Fernando Jaureguizar, Luis Salgado, and Narciso García

Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid,  
28040, Madrid, Spain  
{cda,fjn,L.Salgado,narciso}@gti.ssr.upm.es  
<http://www.gti.ssr.upm.es>

**Abstract.** An efficient automatic detection strategy for aerial moving targets in airborne forward-looking infrared (FLIR) imagery is presented in this paper. Airborne cameras induce a global motion over all objects in the image, that invalidates motion-based segmentation techniques for static cameras. To overcome this drawback, previous works compensate the camera ego-motion. However, this approach is too much dependent on the quality of the ego-motion compensation, tending towards an over-detection. In this work, the proposed strategy estimates a robust motion vector field, free of erroneous vectors. Motion vectors are classified into different independent moving objects, corresponding to background objects and aerial targets. The aerial targets are directly segmented using their associated motion vectors. This detection strategy has a low computational cost, since no compensation process or motion-based technique needs to be applied. Excellent results have been obtained over real FLIR sequences.

## 1 Introduction

Automatic target detection in FLIR imagery are challenging problems due to low signal-to-noise ratio, non-repeatability of target signatures and changes in illumination. Moreover, airborne camera induces a global motion in the sequence (called ego-motion), which produces that static background objects can be detected as moving targets. To overcome the camera ego-motion problem, most of works apply a compensation stage that follows the scheme: computation of the motion vector field, parameter estimation of the global motion and compensation of the global motion [1]-[6]. Each one of these sub-stages has several drawbacks, that as a whole produce a low quality image compensation. Erroneous motion vectors in the motion vector field computation are the most significative drawback, as they can cause an erroneous global motion estimation.

A low quality or erroneous image compensation directly affects to motion-based techniques, that only produces satisfactory results in static images or in perfect compensated images. Besides, these techniques are based on the subtraction of consecutive images [1][2]. Therefore, they usually do not segment

entire moving objects, only some parts of them, due to overlapping of the own objects between consecutive images.

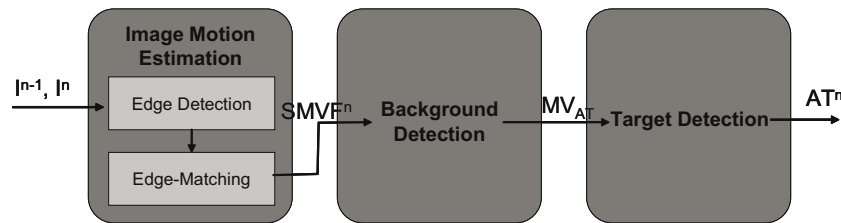
On the other hand, almost all works deal with terrestrial targets. This implies that the airborne camera aims to high-textured earth regions. Nevertheless, this work addresses the aerial target detection, as in [3] and [7]. Consequently, the camera aims to low-textured sky regions, which are not valid to estimate motion due to aperture problem [8]. In this case, the camera ego-motion compensation depends on a reduced set of high-textured cloud and earth regions (if they exist), decreasing its quality.

In this paper a new aerial target detection strategy is presented, which is able to detect moving aerial targets in low-textured sky sequences affected by camera ego-motion. This is achieved by computing a free-error motion vector field, in which high-textured regions are only considered. The motion vector field is analyzed to classify their motion vectors as belonging to background or aerial target regions. Aerial targets are morphologically segmented using the previous motion vector classification. As a result, an accurate and low complexity target detection is obtained, as no static-camera-oriented motion-based technique is applied.

This paper is organized as follows: Section 2 presents an overview of the proposed strategy. Section 3 describes the robust image motion estimation. The background and aerial target detection are presented in Section 4 and 5, respectively. Section 6 shows experimental results obtained over real FLIR sequences. Finally, conclusions are presented in Section 7.

## 2 Strategy Overview

The proposed detection strategy is carried out into three different stages, as shown in Fig. 1. The *Image Motion Estimation* stage automatically detects the edge regions in two consecutive images ( $I^{n-1}$  and  $I^n$ ) of the FLIR sequence. A free-error sparse motion vector field is computed ( $SMVF^n$ ), using only those image regions where were detected edges. The *Background Detection* stage analyzes  $SMVF^n$  to find out if a set motion vector corresponding to background objects exists. If so, those motion vectors are discarded and the rest ( $MV_{AT}$ ) are classified as belonging to aerial targets. The *Target Detection* stage segments all the aerial targets ( $AT^n$ ) presented in  $I^n$ , by morphologically processing those edge regions corresponding to  $MV_{AT}$ .



**Fig. 1.** Stages of the proposed detection strategy

### 3 Image Motion Estimation

This stage detects the edges of a pair of consecutive images,  $I^{n-1}$  and  $I^n$ , and performs an edge-matching to compute a motion vector field, which represents the local motion in the image.

#### 3.1 Edge Detection

A Laplacian of Gaussian based edge detector along with an automatic thresholding is used to detect all the relevant edges in a pair of consecutive images,  $I^{n-1}$  and  $I^n$ .

A Laplacian of Gaussian filter  $LoG$  is applied to the image  $I^n$  to stand out those regions with high intensity variation. As a result,  $I_{LoG}^n$  is obtained, whose intensity values follow a Laplacian distribution, assuming an additive Gaussian noise in the image [11],[12]. The expression of the Laplacian distribution is given by (1):

$$L(x) = \frac{1}{2b} \cdot e^{-(|x-\mu|/b)} \quad (1)$$

where  $\mu$  is the mean and  $b$  is a scale parameter. These parameters are estimated through a robust parameter estimation technique composed by two parts. In the first part, a preliminary estimation,  $\mu_p$  and  $b_p$ , is carried out through the maximum likelihood parameter estimation algorithm. In the second part,  $\mu$  and  $b$  are obtained using the same technique but only over a range of values of  $I_{LoG}^n$ , determined by (2):

$$-4\sigma < I_{LoG}^n < 4\sigma \quad (2)$$

where  $\sigma = 2b_p^2$  is the variance of a Laplace distribution with a scale parameter equal to  $b_p$ .

An adaptive threshold  $T_{LoG}$  is computed from  $\mu$  and  $b$  as in (3):

$$T_{LoG} = \mu - b \cdot \ln \left( 1 - 2 \left| \frac{P_f}{2} - 0.5 \right| \right) \quad (3)$$

where  $P_f$  is the acceptable proportion of false edges (a high value will produce more false edges but detect more true ones, and viceversa).

The intensity values of  $I_{LoG}^n$  smaller than  $T_{LoG}$  are set to zero, obtaining  $I_{ThLoG}^n$ . Then, a zero-crossing technique is applied to  $I_{ThLoG}^n$  to obtain a binary edge image  $E^n$ , which contains all relevant edges. This process is also applied to  $I^{n-1}$  to obtain the edge image  $E^{n-1}$ .

Fig. 2 shows the automatic edge detection process. The FLIR image presented in Fig. 2(a) is filtered by LoG. The intensity distribution of the resulting filtered image is fitted by a Laplacian distribution, as shown Fig. 2(b). An optimum threshold is computed from the parameters of the previously fitted Laplacian distribution. Applying this threshold, the edge image is obtained (Fig. 2(c)). As can be observed, this edge image contains the main edges in the FLIR image, while correctly rejecting those intensity variations due to the noise.

### 3.2 Edge Matching

$E^{n-1}$  and  $E^n$  are morphological dilated with a squared structuring element of size  $5 \times 5$  (which is an acceptable size for selecting the own edge and its more significant neighborhood, that will use to find the edge-based correspondences), obtaining  $DE^{n-1}$  and  $DE^n$ . The dilated edge pixels from  $DE^{n-1}$  are divided into a set of  $k_c$  clusters,  $C_{DE}^{n-1}$ . These clusters are calculated by means of a k-means algorithm, which uses the spatial coordinates of the dilated edge pixels as a feature vector. The number of clusters  $k_c$  is computed as in (4):

$$k_c = \frac{N_{DE}}{N_{pix}} \quad (4)$$

where  $N_{DE}$  is the number of pixels corresponding to the dilated edges, and  $N_{pix}$  is a predefined average number of pixels for each cluster. A high value of  $N_{pix}$  will produce a better quality correspondence but less resolution in the generated motion vector field, and viceversa.

Fig. 3 depicts the clustering of the dilated edge regions  $DE^{n-1}$ , obtained by applying a k-means algorithm over the coordinates of dilated edge regions.

Each cluster in  $C_{DE}^{n-1}$  is composed by a set of pixel coordinates that are used to form clusters of pixels in  $I^{n-1}$ , whose set is denominated  $C_I^{n-1}$ .

The Edge-Matching sub-stage (called in this way because each cluster of  $C^{n-1}$  are formed by pixels belonging to edge regions) compares each cluster of  $C_I^{n-1}$  with the corresponding regions in  $I^n$  (using the same cluster shape), and its adjacent neighborhood located inside a predefined search area  $S_a$ . The search area  $S_a$  is constrained to the dilated edge pixels of  $DE^n$ , since the best correspondence should be another edge region.

The best matching is computed minimizing the mean absolute difference cost function (MAD), whose expression is given in (5):

$$MAD(d_x, d_y) = \frac{1}{N_{pc}} \sum_{(x,y) \in C_{I,i}^{n-1}} |I^{n-1}(x, y) - I^n(x + d_x, y + d_y)| \quad (5)$$

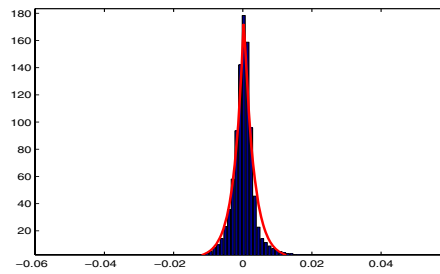
where  $C_{I,i}^{n-1}$  is the cluster  $i$  of  $C_I^{n-1}$  of size  $N_{pc}$  pixels; and  $(d_x, d_y)$  are the coordinates of each candidate motion vector inside  $S_a$ .

The best matching produces a motion vector that defines the movement of one cluster in  $I^{n-1}$  with the corresponding one in  $I^n$ . The set of estimated motion vectors, related to all the clusters of  $I^{n-1}$ , forms a sparse forward motion vector field,  $SFMVF^n$ .

Erroneous vectors can be obtained in  $SFMVF^n$  due to aperture problem [8], the low signal-to-noise ratio of FLIR images and objects that appear or disappear between consecutive images. To discard these erroneous vectors (that could be detected as aerial targets), each motion vector in  $SFMVF^n$  is analyzed. This analysis consists in computing the sparse backward motion vector field  $SBMVF^n$  between  $I^n$  and  $I^{n-1}$ , following the same procedure as for computing  $SFMVF^n$ , but now the clusters of  $I^n$  are those resulting from the best matching in the forward motion estimation process and the search area is constrained



(a)



(b)



(c)

**Fig. 2.** (a) Original FLIR image, (b) Laplacian fitting of the LoG-filtered image intensity distribution, (c) detected edges using the threshold computed through estimated Laplacian parameters

by  $DE^{n-1}$ , the dilated edges of  $E^{n-1}$ . Then, the coherency between forward and backward motion vector fields is verified, by imposing that each couple of associated vectors must satisfy (6):

$$\left( d_x^{SFMVF^n}, d_y^{SFMVF^n} \right) = - \left( d_x^{SBMVVF^n}, d_y^{SBMVVF^n} \right) \quad (6)$$

As a result, an accurate sparse motion vector field  $SMVF^n$  is obtained, free of erroneous motion vectors.



Fig. 3. Clustering of the dilated edge regions using a k-means algorithm

#### 4 Background Detection

The purpose of this stage is to determine if background objects, mainly earth and cloud regions, appear in the FLIR image, and if so, to detect them. The presence of background objects is based on the evaluation of two conditions: quantity and majority conditions. First, the quantity condition is evaluated, which consists in checking if the number of motion vectors in  $SMVF^n$  is larger than a predetermined threshold. Since the background object size is significantly larger than the target size, the number of motion vectors in  $SMVF^n$  in presence of background objects will be much larger than in presence of only aerial targets. If the quantity condition is fulfilled, then the majority condition is evaluated. This condition establishes that at least 50% of motion vectors must follow a coherent motion (corresponding to camera ego-motion). This avoids that aerial targets to be considered as background objects, in the rare situation that an image composed by numerous aerial targets have passed the quantity condition. Notice that the coherent motion corresponding to the background objects can have a magnitude different from zero, even though the background objects are actually static, due to the ego-motion induced by the airborne camera. The coherent motion is modeled through a restricted-affine transformation,  $RAT$ . This transformation is adequate, as the long distance between the camera and both target and background objects allows to simplify the projective camera model into an orthogonal one [7]. The  $RAT$  only considers translations, rotations and zooms, as shown in (7):

$$\begin{bmatrix} x^{n-1} \\ y^{n-1} \\ 1 \end{bmatrix} = \begin{bmatrix} s \cdot \cos \theta & s \cdot \sin \theta & t_x \\ -s \cdot \sin \theta & s \cdot \cos \theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x^n \\ y^n \\ 1 \end{bmatrix} \tag{7}$$

where  $s$ ,  $\theta$ ,  $t_x$  and  $t_y$  are respectively zoom, angle of rotation, horizontal translation and vertical translation; and,  $x^{n-1}$ ,  $y^{n-1}$ ,  $x^n$ ,  $y^n$  are the coordinates

of a determined pixel in  $I^{n-1}$  and  $I^n$  respectively, which are related by the *RAT* transformation.

The *RAT* parameters are estimated by means of a robust parameter estimation technique, based on RANSAC [9], Least Median Square [9] and Median Absolute Deviation algorithms [10]. This estimation technique starts randomly sampling  $S$  pairs of motion vectors from  $SMVF^n$ .  $S$  is calculated to ensure with a probability  $P_s$  that at least one pair of motion vectors is free of outliers (a high value of  $P_s$  will produce a better estimation but more computations, and viceversa). Its expression is given by (8):

$$S = \frac{\log(1 - P_s)}{\log[1 - (1 - \varepsilon^2)]} \quad (8)$$

where  $\varepsilon$  is the expected maximum fraction of outliers in  $SMVF^n$ .

For each pair of motion vectors,  $P_{mv}$ , *RAT* parameters are estimated by solving the equation system presented in (7). The squared residual distance  $r_i^2$  is calculated between each motion vector of  $SMVF^n$  and those obtained from the estimated *RAT* parameters. Then, the median of all  $r_i^2$  is computed, which is used as a quality of goodness of each *RAT* parameter estimation. Therefore, the best fitting  $\widehat{RAT}_e$  is the *RAT* parameter estimation with the minimum value of the median.

The set of inliers vectors  $S_{in}$  is determined through the Median Absolute Deviation algorithm [10]. This uses the set of  $r_i^2$  related to  $\widehat{RAT}_e$  to calculate  $S_{in}$  as in (9):

$$S_{in} = \left\{ mv_i \in SMVF^n \mid (r_i^2) < (2.5 \cdot \hat{\beta})^2 \right\} \quad (9)$$

where  $mv_i$  is a motion vector from  $SMVF^n$  that has associated the squared residual distance  $r_i^2$ , and  $\hat{\beta}$  is the inliers scale estimator given by (10):

$$\hat{\beta} = 1.4826 \cdot \left( 1 + \frac{5}{(N_{mv}) - 2} \right) \cdot \sqrt{\text{median} \{r_i^2\}} \quad (10)$$

where  $N_{mv}$  is the total number of motion vectors in  $SMVF^n$ .

The majority condition is passed if the cardinal of  $S_{in}$  is equal to or larger than  $\frac{N_{mv}}{2}$ , and if so, the members of  $S_{in}$  correspond to background objects. On the contrary, all motion vectors in  $SMVF^n$  will correspond to one or more aerial targets.

## 5 Target Detection

This stage detects aerial targets using the set of motion vectors related to aerial targets,  $S_{AT}$ . If background detection fails,  $S_{AT}$  is set to  $SMVF^n$ . On the contrary, if background detection successes  $S_{AT}$  is set to  $SMVF^n - S_{in}$ , which represents the set of outliers motion vectors in the previous inliers scale estimation process.

The edge regions associated to the members of  $S_{AT}$  are processed by means of a morphological close, using as structuring element an square of size  $D \times D$ , where  $D$  is the mean size of an aerial target. As a result, a set of one or more connected regions is obtained, each one representing a different aerial target.

## 6 Results

The system has been tested with real FLIR sequences captured by an interlaced gray-level infrared camera in 8-12  $\mu\text{m}$  range with a resolution of  $512 \times 512$  pixels. For all the tested sequences one field per frame was selected, therefore the image aspect ratio was modified to 1:2. The camera was mounted on a moving platform that produced a global motion in the sequences. These sequences are mainly composed by low-textured sky regions, and only in some frames by reduced cloud and earth regions. In addition, the sequences are affected by varying illumination conditions.

Fig. 4 shows the motion vector field estimation process, accomplished in the *Image Motion Estimation* stage. Fig. 4(a) shows the original FLIR image with one aerial target and some cloud and earth regions. Fig. 4(b) presents the forward sparse motion vector field, computed between  $I^{n-1}$  and  $I^n$ , where  $I^n$  is used to search the best-correspondences with the dilated edge regions of  $I^{n-1}$ . Fig. 4(c) shows the backward sparse motion vector field, computed as in Fig. 4(b) but between  $I^n$  and  $I^{n-1}$ . And Fig. 4(d) presents the free-error  $SMVF^n$ , composed by those motion vectors of Fig. 4(b) that are coherent with motion vectors of Fig. 4(c), i.e. have the same module but opposite directions. As can be observed, Fig. 4(b), (c) and (d) are sparse motion vector field, since only those regions detected as edges in Section 3.1 are used in the image motion estimation. Notice that some motion vectors from Fig. 4(b) have been discarded in Fig. 4(d), what corresponds with those regions that have appeared/disappeared between consecutive images due to camera ego-motion; or regions that suffer the aperture problem [8], and therefore have a low reliability.

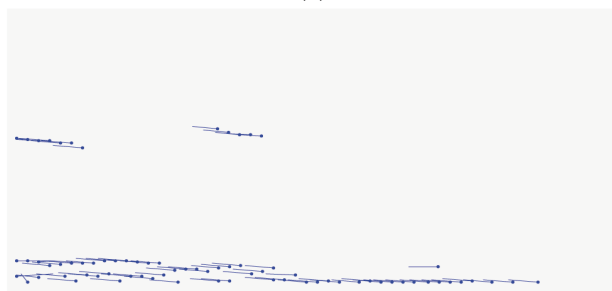
Fig. 5 depicts the aerial target detection process. The  $SMVF^n$  of Fig. 4(d), resulting from the *Image Motion Estimation* stage, is analyzed to detect motion vectors belonging to background or aerial target regions, as shown in Fig. 5(a) (background and aerial target motion vectors are enclosed by a discontinued rectangle and a discontinued oval respectively). Only aerial target motion vectors are morphological processed to segment aerial targets. In this case, the only aerial target is satisfactory segmented, as shown in Fig. 5(b).

Fig. 6 shows another example of the aerial target detection, but with two different aerial target and without any background regions, as shown in Fig. 6(a) (the image has been cropped around the aerial targets to show the process with more clarity). Therefore, the analysis of the corresponding  $SMVF^n$  classifies both connected regions as belonging to aerial target regions (Fig. 6(b), as in Fig. 5(a) aerial target motion vectors are enclosed by discontinued ovals), since the background presence conditions (4) have not been passed. Finally, both aerial targets are segmented through morphological operations, as shown in Fig. 6(c).

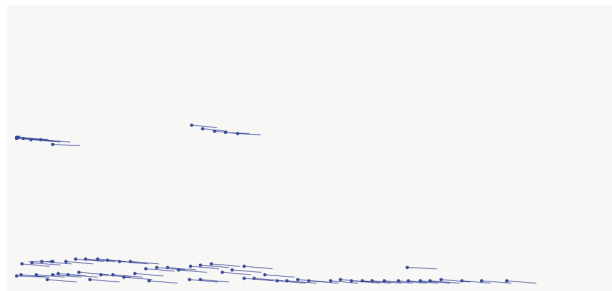




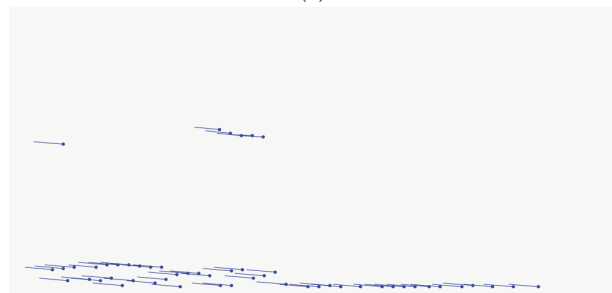
(a)



(b)

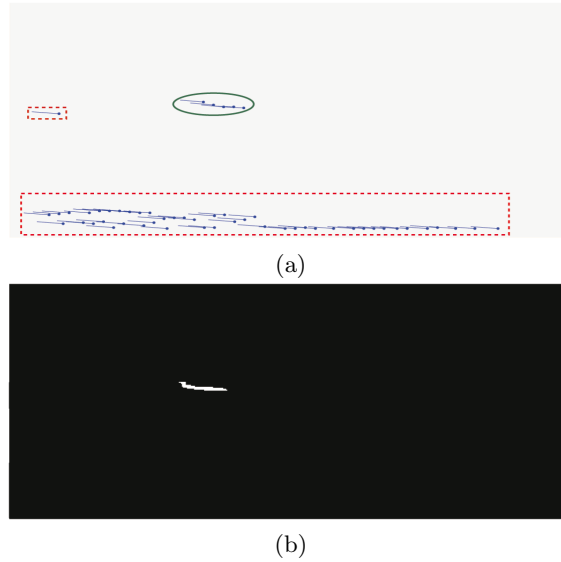


(c)

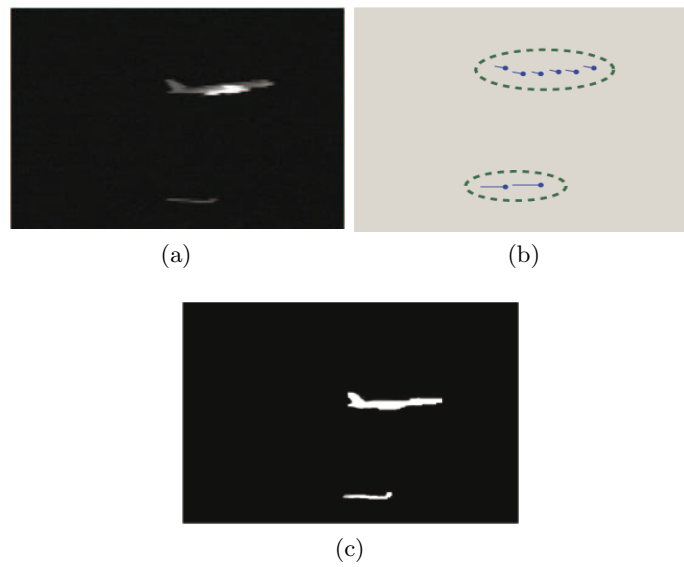


(d)

**Fig. 4.** (a) Original FLIR image, (b) forward sparse motion vector field, (c) backward sparse motion vector field and (d) free-error *SMVF*



**Fig. 5.** (a) Motion vector classification of Fig 4(d) into background objects and aerial targets; (b) Aerial target segmentation, obtained by the morphological processing of the regions associated to aerial target motion vectors



**Fig. 6.** (a) The cropped original FLIR image containing two aerial targets; (b) motion vector classification of the  $SMVF^n$ , obtained from (a) and the previous image in the sequence; (c) morphological segmentation of the two aerial targets presented in (a), using the motion vectors from (b)

The proposed target detection is efficient with targets of reduced size. However, when the target area is less than 50 pixels, its performance begins decreasing.

The entire set of FLIR sequences has been processed, obtaining an average detection rate of 98.2% and an average false alarm rate of 3.8%, using the following parameters:  $P_f = \frac{1}{512 \times 512}$ ,  $N_{pix} = 256$ ,  $\epsilon = 0.4$  and  $P_s = 0.9999$ . These results demonstrate the excellent performance of this detection strategy.

## 7 Conclusions

A novel strategy for detecting aerial moving targets in airborne FLIR imagery has been presented in this paper. Instead of compensating the camera ego-motion for applying static-oriented motion-based techniques, the proposed strategy directly analyzes the image motion (calculated as a motion vector field) to separately cluster background and aerial target regions. The aerial targets are segmented by morphologically processing the aerial target regions. In order to achieve this detection, the computation of a free-error motion vector field is required. This is accomplished by the combination of two strategies: using only the edge regions to compute the motion vectors; and testing the coherency of the motion vectors belonging to the forward and backward motion vector fields. In addition to the gained reliability, a low complexity is achieved, since only a reduced set of image regions is processed. The results presented in Section 6 demonstrate the high efficiency of this detection strategy, which is able to accurately detect multiple aerial targets under ego-motion and clutter conditions.

## Acknowledgements

This work has been partially supported by the Ministerio de Ciencia y Tecnología of the Spanish Government under project TIN2004-07860 (Medusa) and by the Comunidad de Madrid under project P-TIC-0223-0505 (Pro-Multidis).

## References

1. Strehl, A., Aggarwal, J.K.: Detecting moving objects in airborne forward looking infra-red sequences. In: Proc. IEEE Workshop on Computer Vision Beyond Visible Spectrum, pp. 3–12. IEEE Computer Society Press, Los Alamitos (1999)
2. Strehl, A., Aggarwal, J.K., MODEEP: a Motion-Based Object Detection and Pose Estimation Method for Airborne FLIR Sequences. *Machine Vision and Applications*. 11(6), 267–276 (2000)
3. Estalayo, E., Salgado, L., Jaureguizar, F., García, N.: Efficient image stabilization and automatic target detection in aerial FLIR sequences. *Automatic Target Recognition XVI*. In: Proc. of the SPIE, vol. 6234 (2006)
4. Seok, H.D., Lyou, J.: Digital Image Stabilization using Simple Estimation of the Rotational and Translational Motion. *Acquisition, Tracking and Pointing XIX*. Proc. of SPIE 5810, 170–181 (2005)

5. Yilmaz, A., Shafique, K., Lobo, N., Li, X., Olson, T., Shah, M.A.: Target-tracking in FLIR imagery using mean-shift and global motion compensation. In: Proc. IEEE Workshop Computer Vision Beyond Visible Spectrum, IEEE Computer Society Press, Los Alamitos (2001)
6. Yilmaz, A., Shafique, K., Shah, M.: Target Tracking in Airborne Forward Looking Infrared Imagery. *Image and Vision Computing Journal*. 21(7), 623–635 (2000)
7. Meier, W., Stein, H.: Estimation of object and sensor motion in infrared image sequences. In: Proc. IEEE. Int. Conf. on Image Processing, vol. 1, pp. 568–572. IEEE, Los Alamitos (1994)
8. Wechsler, H., Duric, Z., Fayin, L., Cherkassky, V.: Motion estimation using statistical learning theory. *IEEE Trans. on Pattern Analysis and Machine Intelligence*. 26(4), 466–478 (2004)
9. Stewart, C.V.: Robust parameter estimation in computer vision. *SIAM Reviews*. 41(3), 513–537 (1999)
10. Meer, P., Stewart, C.V., Tyler, D.: Robust computer vision: an interdisciplinary challenge. *Computer Vision and Image Understanding*. 78(1), 1–7 (2000)
11. Rosin, P.: Edges: Saliency measures and automatic thresholding. *Machine Vision and Applications*. 9(4), 139–159 (1999)
12. Srivastava, A., Lee, A.B., Simoncelli, E.P., Zhu, S.C.: On Advances in Statistical Modeling of Natural Images. *Journal of Mathematical Imaging and Vision*. 18, 17–33 (2003)