

TARGET DETECTION THROUGH ROBUST MOTION SEGMENTATION AND TRACKING RESTRICTIONS IN AERIAL FLIR IMAGES

Carlos R. del-Blanco, Fernando Jaureguizar, Luis Salgado, and Narciso García
Grupo de Tratamiento de Imágenes
Universidad Politécnica de Madrid - 28040 Madrid (Spain)
{cda,fjn,lsa,narciso}@gti.ssr.upm.es

ABSTRACT

An efficient automatic moving target detection and tracking system in airborne forward looking infrared (FLIR) imagery is presented in this paper. Due to camera ego-motion, these detection and tracking tasks are challenging problems. Besides, previously proposed techniques are not suitable for aerial images, as the predominant regions are non-textured. The proposed system efficiently estimates not only the camera motion but also the target motion, by means of an accurate motion vector field computation and robust motion parameters estimation technique. This information allows accurately to segment each target, and tracking them with ego-motion compensation. Verification of tracking restrictions helps detecting true targets while reducing very significantly the false alarm rate. Excellent results have been obtained over real FLIR sequences.

Index Terms— motion segmentation, robust motion estimation, ego-motion compensation, trajectory restrictions, FLIR images.

1. INTRODUCTION

Automatic moving target detection and tracking in forward looking infrared (FLIR) imagery are challenging problems due to low signal-to-noise ratio, non-repeatability of target signatures, and illumination variation. These produce random variations of the image gray level values. Besides, in airborne applications, the sequence is affected by camera ego-motion, that makes detection and tracking processes extremely difficult. Therefore, accurate image stabilization strategies are required, which additionally will allow consistent trajectories computation.

Nevertheless, ego-motion estimation for image stabilization is rarely considered in FLIR target detection and tracking research works, and most of them are focused on air-to-earth scenarios. Some works [1][2] incorporate

ego-motion compensation for image stabilization to compute target detection through image differencing. However, these segmentation techniques offer poor results because of the aforementioned drawbacks of FLIR images together with the reduced size of the targets. Others, [3][4], use ego-motion compensation for target tracking and not for target detection, and its computation is restricted only to those cases where tracking fails due to large camera displacements. Therefore, although tracking is achieved in these cases, there is a lack of temporal consistency in the computed target trajectories as most of them are affected by non-compensated ego-motion.

Here, the objective is the accurate and efficient detection and tracking of aerial targets in FLIR sequences composed by large uniform sky regions and reduced cloud or earth regions. Therefore, ego-motion estimation is particularly difficult as no stable reference background can be found due to the lack of large textured background areas. In [5] aerial target scenarios are addressed, looking for low computational cost approaches. Nevertheless, that technique could be biased when predominant non-textured regions are present in the scene.

In this paper a new target detection and tracking strategy is proposed, which detects small size targets in predominant uniformed sky sequences affected by camera ego-motion. A motion analysis, based on robust parameters estimation, allows to segment potential moving targets and tracking them without ego-motion distortion. Tracking restrictions to efficiently reduce the number of false alarms are applied.

This paper is organized as follows: Section 2 presents an overview of the proposed strategy. Section 3 and 4 introduce local motion estimation and motion analysis approaches, while Section 5 describes the tracking restrictions and their verification. Excellent experimental results obtained over real FLIR sequences and conclusions are presented in Section 6 and 7.

2. SYSTEM OVERVIEW

The system, shown in Figure 1, is composed by three processing modules: Local Motion Estimation, Motion Analysis and Tracking Restrictions Verification.

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

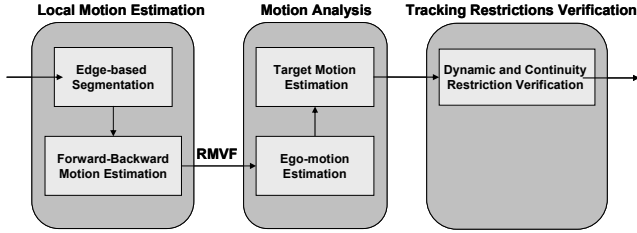


Figure 1: System block diagram.

Consecutive sequence images, I_{n-1} and I_n , are processed by the Local Motion Estimation module to estimate a robust motion vector field, $RMVF$, through a combined forward and backward efficient block-based motion estimation strategy restricted to edge-based regions. $RMVF$ is analyzed to robustly estimate camera and target motion parameters, assuming an affine motion model. This information is used to segment potential moving targets and perform their tracking removing the ego-motion, TR_n^i . Over this tracking information, restrictions are applied to detect true targets, DT_n , and discard moving clutter in the Tracking Restriction Verification module.

3. LOCAL MOTION ESTIMATION

A motion vector field is computed through a block-based forward-backward motion estimation strategy. Due to the aperture problem [6], which arises from non-textured regions, motion estimation is restricted to blocks centered on significant image edges.

3.1. Edge image generation

Edge detection is carried out by a Canny detector, which requires two thresholds to be fixed to discard those edges either non relevant or produced by the noise present in the image. Here, the noise distribution is characterized as Gaussian, and its variance is obtained from the histogram of the union of the vertical and horizontal gradient images, obtained through the derivative filter $D=1/2[-1 \ 0 \ 1]$. Therefore, upper and lower Canny thresholds are set to $2.5\sigma_n$ and $1.5\sigma_n$, where σ_n is the histogram standard deviation. So, this strategy is automatically adapted to the level of noise present in the images.

Each resulting binary image is morphologically dilated to allow a safety margin in the edge location, obtaining the binary image, E_n . Figure 2 (a) shows an original FLIR image, and (b) the corresponding edge-based binary image.

3.2. Forward-Backward Motion Estimation

A robust motion vector field, $RMVF$, is generated by means of a backward and forward block-based motion estimation restricted to image regions determined by the one-valued pixels in E_{n-1} and E_n .

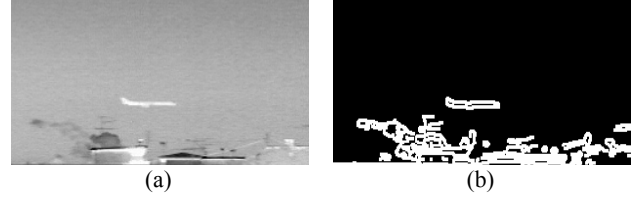


Figure 2: (a) Original image I_n , (b) dilated edge image E_n

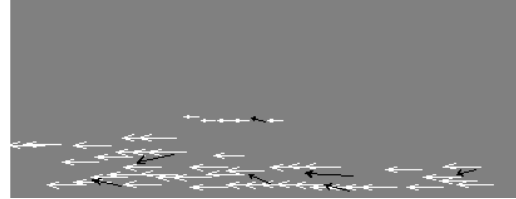


Figure 3: Sparse motion vector field with detected errors.

Forward motion estimation is computed through an adaptive search block matching technique, which generates a sparse motion vector field between I_{n-1} and I_n . Block centers must belong to edge pixels and, to minimize overlapping, the minimum distance between two reference blocks is the block size. So, each block center (\vec{r}_b) of I_{n-1} must be one edge pixel of E_{n-1} and each block center (\vec{r}) of I_n must be one edge pixel of E_n . For each considered block in I_{n-1} , the motion vector is computed as:

$$\vec{d}_1 = (\vec{r}_{m1} - \vec{r}_b) \mid \vec{r}_{m1} = \arg \max_{\vec{r} \in S_b} (NCCF(\vec{r}_b, \vec{r})) \quad (1)$$

where $NCCF(\vec{r}_b, \vec{r})$ is the Normalized Cross Correlation Function between the block of I_{n-1} centered in \vec{r}_b and another block of I_n centered in \vec{r} . \vec{r} belongs to the search area, S_b , restricted to a circle of radius R centered in \vec{r}_b .

Next, the backward motion vector field between images I_n and I_{n-1} is computed like the forward estimation, except that now the blocks of I_n are centered in \vec{r}_{m1} (those pixels resulting from the forward motion estimation process as in (1)). As a result, motion vector is computed as:

$$\vec{d}_2 = (\vec{r}_{m2} - \vec{r}_{m1}) \mid \vec{r}_{m2} = \arg \max_{\vec{r} \in S_b} (NCCF(\vec{r}_{m1}, \vec{r})) \quad (2)$$

Finally, the coherency between forward and backward motion vector fields is verified by imposing that vectors belonging to $RMVF$ must satisfy $d_1 = -d_2$, which removes low reliability motion vectors. Figure 3 presents the sparse $RMVF$ (in white) corresponding to the image in Figure 2 (a), where removed erroneous motion vectors are shown in black.

4. MOTION ANALYSIS

$RMVF$ is analyzed to accurately segment the different independent moving regions, IMR , whose motion is modeled through a restricted affine transformation, RAT .

This transformation is adequate, as the long distance between the camera and the moving objects allows to consider orthographic projection.

The RAT parameters associated to the camera ego-motion, RAT_{cam} , are accurately estimated through a robust estimation strategy, under the assumption that the motion vectors corresponding to the targets are outliers (their associated regions are a minority with respect to the regions associated to textured background in the considered scenarios). Next, outlier motion vectors are clustered according to a connectivity criterion of their associated blocks, and then, the same robust motion estimation strategy is applied to segment the potential targets.

The tracking of each potential target is performed by compensating the camera ego-motion, RAT_{cam} , in each target motion, defined by its own RAT_{tar}^i .

4.1. Ego-motion Estimation

Camera ego-motion is modeled by a RAT that is defined by the matrix:

$$RAT = \begin{bmatrix} s \cdot \cos \theta & s \cdot \sin \theta & t_x \\ -s \cdot \sin \theta & s \cdot \cos \theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

where s is the scale, θ is the angle of rotation and the pair (t_x, t_y) represents the translations along the axes. These are estimated through a robust parameter estimation technique presented in [7], based on RANSAC and Least Median Square algorithms. S subsets of $RMVF$, each one formed by k vectors, are randomly sampled, where S is computed as:

$$S = \frac{\log(1 - P)}{\log(1 - (1 - \varepsilon^k))} \quad (4)$$

which can be interpreted as the minimum number of subsets that ensures with a probability P to find at least one subset without outliers. ε is the maximum fraction of outliers contained in $RMVF$, and k is the dimension of the parameter space.

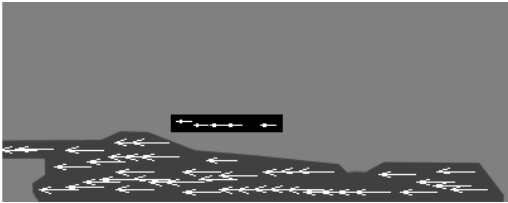


Figure 4: motion vector segmentation.

For each subset (ψ in $RMVF$) RAT parameters (RAT_ψ) are estimated through Least Mean Square (LMS), and then, they are tested as a hypothesized fitting to the whole $RMVF$ by calculating the median of the squared residual distance

(r_{i,RAT_ψ}^2) of the motion vectors $d_i \in \{RMVF - \psi\}$. The best fitting, \hat{RAT}_ψ , to $RMVF$ is given by the expression:

$$\hat{RAT}_\psi = \arg \min_{RAT_\psi} \left(\text{median}_{d_i \in \{RMVF - \psi\}} \{r_{i,RAT_\psi}^2\} \right) \quad (5)$$

However, \hat{RAT}_ψ suffers a lack of accuracy since only a limited subset of $RMVF$ has been used for the estimation. The Median Absolute Deviation algorithm [8] overcomes this drawback using the entire set of inliers vectors, d_1^{in} , in the fitting process, which is determined by the expression:

$$d_1^{in} = \left\{ d_1 \mid r_{i,RAT_\psi}^2 < (2.5 \cdot \hat{\beta})^2 \right\} \quad (6)$$

where $\hat{\beta}$ is the inliers scale estimator given by:

$$\hat{\beta} = 1.4826 \cdot \left(1 + \frac{5}{N - k} \right) \cdot \sqrt{\text{median}_{d_i \in \{RMVF - \psi\}} \{r_{i,RAT_\psi}^2\}} \quad (7)$$

where N is the total number of motion vectors in $RMVF$. Finally, the camera ego-motion RAT parameters, RAT_{cam} , are obtained by applying LMS over d_1^{in} .

4.2. Target Motion Estimation

Outlier motion vectors, $d_1^{out} = \{RMVF - d_1^{in}\}$, are used to estimate the target RAT parameters, RAT_{tar}^i . d_1^{out} are clustered according to the spatial connectivity of their corresponding blocks. Each cluster is a potential target whose RAT_{tar}^i is estimated as described in the previous section for the RAT_{cam} estimation. Potential targets are segmented considering only the blocks associated to the inliers motion vectors used to compute each RAT_{tar}^i .

Considering \bar{TR}_n^i and \bar{TR}_{n-1}^i as the target positions at instants n and $n-1$, the evolution of the target positions (the target tracking) with the camera ego-motion compensated is computed as:

$$\bar{TR}_n^i = RAT_{cam}^{-1} \left(RAT_{tar}^i \cdot \bar{TR}_{n-1}^i \right) \quad (8)$$

Figure 4 shows the resulting background (in grey) and target (in black) motion based segmentation corresponding to the $RMVF$ presented in Figure 3.

5. TRACKING RESTRICTION VERIFICATION

False targets are discarded by imposing continuity and motion-size restrictions to each potential target tracking.

Birds, which are potential moving targets, are removed by imposing a motion-size restriction that considers that target regions must overlap between consecutive images, while those regions corresponding to birds will never overlap if their size is similar to the targets.

Continuity restriction discards potential targets that only appear in a limited number of consecutive frames, due to

random gray level variations, typical from FLIR images as described in the introduction. Therefore, the trajectory of each potential target is defined as:

$$\bar{T}J_n^i = \{\bar{T}R_m^i \dots \bar{T}R_{n-1}^i, \bar{T}R_n^i\} \quad (9)$$

where m is the instant in which the potential target is detected for the first time in the scene. A potential target is removed if $|\bar{T}J_n^i| < th$, where $|\cdot|$ is the cardinal, and th is a scalar threshold.

As a result of applying both restrictions, the false alarm rate is significantly reduced, and true targets, DT_n , are obtained.

6. EXPERIMENTAL RESULTS

The system has been tested with real FLIR sequences captured by interlaced 8-bit-gray-level infrared camera with resolution 512x512 pixels, mounted on a moving platform. The sequences are composed by non-textured sky regions (that predominate over cloud or earth ones), targets of different size and birds. In addition, they are affected by camera ego-motion and varying illumination.

Figure 5 shows three frames of two different sequences. The upper ones show urban clutter and a saturated target which is accurately segmented and tracked. The three lower ones have country clutter and a contrasted target which has been also correctly detected discarding some small regions corresponding to birds.

The performance of the system has been measured by means of the Detection Rate and False Alarm Rate with and without the tracking restriction verification phase. An excellent Detection Rate of 99.4% is achieved, and the False Alarm Rate of the complete system is 0.8%. The efficiency of the proposed strategy is confirmed by the sharp reduction of the False Alarm Rate that is 11.3% before applying the tracking restriction verification.

7. CONCLUSIONS

An efficient automatic moving target detection and tracking system has been presented, which accurately detects moving targets in low-textured airborne FLIR images. Image motion is analyzed through a robust motion parameters estimation technique, which allows accurately to segment accurately potential moving targets and tracking them. Tracking restrictions, applied over the ego-compensated trajectories, are verified to identify true targets while the false alarms are drastically reduced. Exceptional results have been obtained confirming the efficiency and reliability of the proposed system.

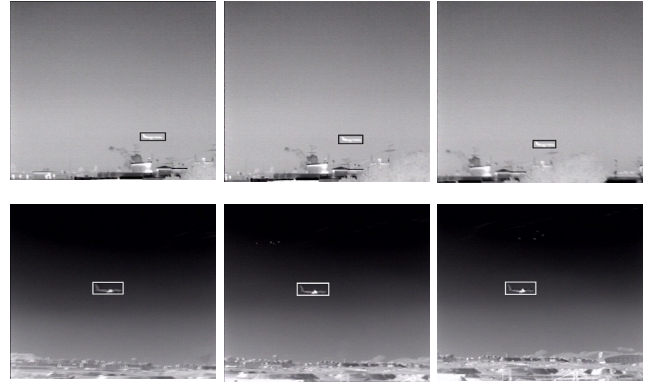


Figure 5: Target detection and tracking.

8. REFERENCES

- [1] A. Strehl, J. K. Aggarwal. Detecting moving objects in airborne forward looking infra-red sequences. *Proc. IEEE Workshop on Computer Vision Beyond Visible Spectrum*, Fort Collins, pp. 3–12, 1999.
- [2] A. Strehl, J.K. Aggarwal, “MODEEP: a Motion-Based Object Detection and Pose Estimation Method for Airborne FLIR Sequences”, *Machine Vision and Applications*, vol. 11, no. 6, pp. 267–276, 2000.
- [3] A. Yilmaz, K. Shafique, N. Lobo, X. Li, T. Olson, M. A. Shah, “Target-tracking in FLIR imagery using mean-shift and global motion compensation,” *Proc. IEEE Workshop Computer Vision Beyond Visible Spectrum*, Kauai, HI, 2001.
- [4] A. Yilmaz, K. Shafique, M. Shah "Target Tracking in Airborne Forward Looking Infrared Imagery," *Image and Vision Computing Journal*, vol. 21, no. 7, pp. 623–635, 2003.
- [5] E. Estalayo, L. Salgado, F. Jaureguizar, N. García, “Efficient image stabilization and automatic target detection in aerial FLIR sequences”, *Automatic Target Recognition XVI, Proc. of the SPIE*, vol. 6234, paper 62340N, 2006.
- [6] H. Wechsler, Z. Duric, L. Fayin, V. Cherkassky, “Motion estimation using statistical learning theory”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 4, pp. 466–478, 2004.
- [7] C.V. Stewart, “Robust parameter estimation in computer vision”, *SIAM Reviews*, vol. 41, no. 3, pp. 513–537, 1999.
- [8] P. Meer, C.V. Stewart, D. Tyler, “Robust computer vision: an interdisciplinary challenge”, *Computer Vision and Image Understanding*, vol. 78, no 1, pp. 1–7, 2000.