A Robust Real-time Image Algorithm for Moving Target Detection from Unmanned Aerial Vehicles (UAV)

Mathieu Pouzet[1,2], Patrick Bonnin[1,2], Jean Laneurit[1] and Cedric Tessier[1]
[1]*Effidence, Clermont-Ferrand, France*
[2]*LISV, Versailles, France*

Keywords: Aerial Imagery, Moving Target Detection, Registration, Spatio-temporal Trajectories.

Abstract: We propose a real time method for moving target detection from a camera embedded on a UAV. As the camera is moving, we must estimate the background motion in order to compensate it and then perform the moving target detection. This compensation is realized by an image registration method. For this, we use an hybrid method using global minimization and feature-based approaches, with a pyramidal implementation. The good results obtained for registration give us the potential moving targets. As some wrong detections still appear, due to noise, occlusions or local change of illuminations, we worked on a robust spatio-temporal tracker able to decide if potential targets are real moving targets or not. The algorithm must reach real time performances for VGA images at 30 fps with a standard PC. We have tested our method on different sequences and show the good results obtained thanks to the high precision in the image registration and the spatio-temporal tracker.

1 INTRODUCTION

Among the goals of video surveillance, one is the detection and tracking of all the moving targets in the scene. This is a very difficult task, even more when the camera is moving and that moving targets are difficult to characterize (possible low speed, small size, low contrast ...), as in aerial surveillance scenarios. Possible applications are highway or border surveillance.

This paper focuses on the Moving Target Detection (MTD) problem from UAV. The targets are moving with a low speed and the images taken from the camera embedded on the UAV suffer from an important and irregular movement. Our main objective is to highlight the pixels belonging to the targets of the current image that do not match with the background motion. This latter may be very high compare to the target motion. We do not want to miss small targets with low speed and contrast. It is why background motion estimation with a high accuracy is needed. But we do not forget that target motion may deteriorate the quality of the background motion estimation.

As spurious detections may appear, we propose a second step of filtering realized by a tracker, keeping only moving targets with a spatio-temporal coherent motion. It is able to track with occlusions during a few images The result is given under the form of a Ground Moving Target Indicator (GMTI) map.

In the following of the paper, we first present a state of the art about moving target detection from UAV (sec. 2). Then we propose the outline of the method (sec. 3), and its details : the image registration (sec. 4) and the moving target detection (sec. 5). Finally, we will evaluate our method on simulated and real image sequences (sec. 6) and conclude (sec. 7).

2 RELATED WORK

Lots of approaches have been proposed to address the MTD problem in airborne video analysis. Some algorithms use interest points in image, that have to be labelled as static or dynamic(Richard Evans, 2007)(Cohen and Medioni, 1998) (Rodrguez-Canosa et al., 2012). This approach is the same as we can find in SLAMMOT (Simultaneous Localization And Mapping With Moving Objects Tracking) systems(chih Wang et al., 2004) (Migliore et al., 2009). The main problem of this technique is the fact that features may not be present on moving targets to be tracked, and especially if their size is small and if they have a low contrast with the background.

It is why we prefer to adopt a “kind of”[1] im-

[1] Output of M-Estimator (sec. 4.3.2)

Pouzet M., Bonnin P., Laneurit J. and Tessier C..
A Robust Real-time Image Algorithm for Moving Target Detection from Unmanned Aerial Vehicles (UAV).
DOI: 10.5220/0005052802660273
In *Proceedings of the 11th International Conference on Informatics in Control, Automation and Robotics* (ICINCO-2014), pages 266-273
ISBN: 978-989-758-039-0

age subtraction approach after a global background motion compensation. Assuming that pixels intensities do not vary between two consecutive frames, the method consists in applying a threshold to the difference of the registered images to label pixels as moving or not(Irani and Anandan, 1998)(Ali and Shah, 2005). This technique allows for detecting much more moving targets than feature-matching based approach.

To create robust map, it is important to take spatio-temporal coherency into account. It is why (Odobez and Bouthemy, 1997)(Crivelli et al., 2011) use spatio-temporal Markov Random Field (MRF) to create a statitiscal labelling. We tested it : segmentation results are good but the tuning of parameters is too critical for our application.

In order to obtain a better segmentation of moving targets, some authors propose to create a mozaic image representation of the background. As moving targets do not have to be included in the mozaic, a background model is created with the median of images (Reilly et al., 2010) or mode (Lin et al., 2011), i.e most frequent value in a sample. For computation time reason, the GPU is used. About the tracking method, moving targets are associated into tracklets (Lin et al., 2011) linked by their appearance, needing targets to contain enough information to be tracked correctly, that is not always our case.

Our main contribution is the proposal of a robust real-time MTD algorithm, working in various cases, even for feature-less images or large camera motion. It works even if the targets have a small size or a very low contrast with the background.

3 OUTLINE OF THE METHOD

Our method is composed of two main steps (see Fig.1) :

- Registration of two consecutive images,
- Moving Target Detection (MTD) :
  - Prediction (using two consecutive images): potential target detection D_t by gathering pixels whose motion differ from background motion,
  - Validation (using N previous potential targets $\{D_{t-1}...D_{t-N}\}$ from N previous images): true target detection by filtering with a spatio-temporal tracker

The outputs of the image registration step are :

- the global transformation $\mathscr{T}(\Theta)$ (ie. its parameters) ,
- the weight image of M-Estimator W, indicating the moving target pixels.

Generally, moving targets are detected by a high threshold value on difference between registered images : that eliminates potential moving targets with low contrast with background. Here, the results are obtained using an M-Estimator (Prediction step), adapting the threshold according to the camera noise, followed by a tracker (Validation step). This one is able to follow different kinds of targets, even 3×3 pixels (sec. 5.2).

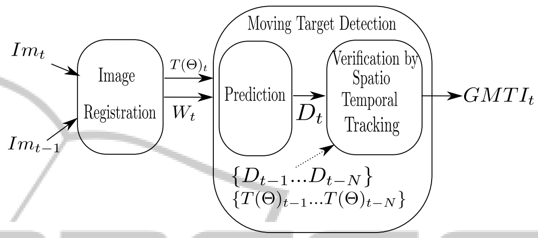


Figure 1: Overall method scheme of Ground Moving Target Indicator (GMTI) using two images at time t and $t-1$. D_t means potential detected targets at time t.

4 IMAGE REGISTRATION

Our objective is to find the transformation $\mathscr{T}(\Theta)$ able to register an image A in the reference frame of the reference image B. Given the medium-altitude of the UAV and therefore our constraint of planarity, the transformation to estimate is a similarity (Hartley and Zisserman, 2004), composed of an homothety of ratio s, a rotation of angle θ, and a translation of vector $\left[t_u, t_v\right]^T$. defined as :

$$\left[u, v, 1\right]_A^T = \mathscr{T}(\Theta)\left[u, v, 1\right]_B^T$$

with

$$\mathscr{T}(\Theta) = \begin{bmatrix} s\,cos\theta & -s\,sin\theta & t_u \\ s\,sin\theta & s\,cos\theta & t_v \\ 0 & 0 & 1 \end{bmatrix}$$

and with parameters vector $\Theta = \left[s, \theta, t_u, t_v\right]^T$. The function $\mathscr{W}$ that maps pixels X of image A in the frame of image B by the use of transformation $\mathscr{T}(\Theta)$ is called a warping function and is such as :

$$\mathscr{W}(\Theta, X) = I(\mathscr{T}(\Theta), X)$$

with I a bilinear interpolation function.

Two main approaches could be considered for image registration: the global methods (sec. 4.1) using a similarity criteria (e.g pixel intensity) between the two images and the geometric methods (sec. 4.2) using the matching of static features (edges, interest points ...) found in the two images.

4.1 Global Methods

These methods estimate the transformation by minimizing or maximizing a criteria (difference of pixel intensity, entropy ...) between each image A and B. The objective is to iteratively get a warped version of image A that corresponds to the reference image B. This could be done by splitting the image in regions (to get the optical flow for instance,(yves Bouguet, 2000)) or more globally by warping the whole image ((Odobez and Bouthemy, 1997) (Benhimane and Malis, 2004),(Dame and Marchand, 2012)(Zitov and Flusser, 2003)...).

To estimate $\mathscr{T}(\Theta)$, classical minimization techniques proceed in an iterative way, with a first proposed estimation (for initialization) $\mathscr{T}(\hat{\Theta}) = Id$, and then by progressively estimating the transformation. By knowing that a multiplication of similarities is still a similarity, at each iteration we get : $\mathscr{T}(\hat{\Theta}) = \mathscr{T}(\hat{\Theta})\mathscr{T}(\Delta_\Theta)$ with Δ_Θ the parameters variation for a given iteration. At convergence, when $\|\Delta_\Theta\|$ tends to 0 , we get : $\mathscr{T}(\Theta) = \mathscr{T}(\hat{\Theta})$. Δ_Θ must be calculated to iteratively estimate $\mathscr{T}(\Theta)$.

If we note $\overline{A}$ and $\overline{B}$ the vectors respectively containing the pixels intensity of image A and B after gray conversion, and $\varepsilon(\Delta_\Theta) = \overline{B} - \overline{A}(\mathscr{W}(\hat{\Theta}, \mathscr{W}(\Delta_\Theta, X))))$, the cost function can be written as :

$$f(\Delta_\Theta) = \frac{1}{2}\|\varepsilon(\Delta_\Theta))\|^2$$

In literature, Gauss-Newton is the most consistently-used method in image processing (KLT Lucas-Kanade (Baker and Matthews, 2002), (yves Bouguet, 2000)). It works perfectly in major cases. To answer our constraint of low-computation time, we prefered the more recent method developed by (Benhimane and Malis, 2004), named ESM (Efficient Second order Minimization).

This methods allows us to estimate $\mathscr{T}(\Theta)$ by the calculation of Δ_Θ such as :

$\Delta_\Theta = -(J_{esm}^T J_{esm})^{-1} J_{esm}^T \varepsilon(0)$ (see Details in (Benhimane and Malis, 2004)).

This method is not sensitive to the initial segmentation quality,contrarily to the geometric method. Its main drawback is its uncapacity to estimate large movements as this is a local technique.

4.2 Geometric Feature-based Methods

The objective of these methods is to segment the images in interest points (Harris and Stephens, 1988)(Smith and Brady, 1995)(Lowe, 2004)(Bay et al., 2008), edges (de Cabrol et al., 2005) or regions (A.Goshtasby, 1986)(K.Poornima, 2012)(de Cabrol et al., 2005). Then it aims at matching these data between the two images in order to calculate their motion.

The main advantages are a low algorithmic complexity due to the low number of information to be extracted and the possibility to estimate large movements of the camera. But the drawback is the necessity to extract a sufficient number of primitives with a good distribution in the image. Most of time, it results in a lower motion estimation precision than with the use of the global method because it uses only a part of the image information.

Regarding the complementarity of these two approaches lead us to the choice of an hybrid method detailed below.

4.3 Proposed Image Registration Method

4.3.1 Robustness to Large Camera Movements

ESM is a local technique, only authorizing small motion between images. This constraint can be minimized using a pyramidal implementation. For a VGA image, a 3 levels pyramid is used, with a gaussian filter of window 5×5: level L_2: 160×120, level L_1: 320×240, level L_0: 640×480.

However, it is not sufficient for certain of our applications. A second way to minimize this constraint is necessary. It consists in coupling the pyramidal ESM with a feature-matching based motion estimation algorithm. Among a lot of feature-matching based methods ((Harris and Stephens, 1988)(Smith and Brady, 1995)(Lowe, 2004)(Bay et al., 2008)), we choose a simple interest points matching technique by Harris corners (Harris and Stephens, 1988) and correlation. For computation constraint, corners found in image A has only to be matched with corners of image B.

This method is both able to estimate large movements and is a quick method. However, it is not able to give a transformation precise enough for our MTD application : Harris corners being not stable enough to rotation and scale. Moreover, nothing can guarantee us that a sufficient number of interest cornes will be found in the image. Therefore an hybrid algorithm is used to initialize the ESM in case of large movements by giving to it a first guessed estimation at a lower pyramid level.

4.3.2 Robustness to Noise, Moving Targets and Occlusions

Both approaches for image registration do not give a correct transformation estimation in presence of noise, occlusions, moving targets in the image that do not respect the global motion of the camera ... It is why we need to robustify these approaches with robust estimators.

Therefore an M-estimator is used at each ESM iteration to modify the weight of each pixel in the final calculation. This estimator is recommended by (Malis and Marchand, 2006)(Klose et al., 2013)(Odobez and Bouthemy, 1997) for robust estimation in real-time vision. This is a very good way in order not to take into account the moving targets in the image, that would deteriorate the global transformation estimation.

It is why for our application, we will use the weight function w as defined by (Huber et al., 1981) :

$$\begin{cases} w(x,y) = 1 & si \quad \varepsilon_N(x,y) \leq C \\ w(x,y) = \frac{C}{\varepsilon_N(x,y)} & si \quad \varepsilon_N(x,y) > C \end{cases} \quad with$$

- $\varepsilon_N(x,y) = \frac{|\varepsilon(x,y) - med(\varepsilon(x,y))|}{\sigma}$,
- $\sigma = 1.4826 \ MAD$, $MAD = med(|\varepsilon(x,y) - med(\varepsilon(x,y))|)$ the "Median Absolute Deviation",
- $C = 1.345$ the constant allowing to get a 95% asymptotic efficiency.

These weights w are calculated at each ESM iteration such as : $\Delta_\Theta = -(J_{esm}^T W J_{esm})^{-1} J_{esm}^T W \varepsilon(0)$ with W the diagonal weight matrix for each pixel.

About feature-based algorithm, we use a modified version of RANSAC algorithm (Fischler and Bolles, 1981) to reject the outliers. This version allows to deal with multi-hypotheses matchings. Rather than taking into account only the best match for a pair of interest point (one in each image) based on correlation criteria, several matches are considered.

Therefore, at each iteration, RANSAC draws an hypothetic match following a cumulative uniform distribution based on correlation scores. Higher is the correlation score between a pair of interest points (one of each image), better is the chance for this corner to be drawn.

To meet computation time constraint, the search for matchings is not performed in the whole image but in an area whose the size is dependent on the estimated motion and its covariance, given from a level of the pyramid to another one.

4.3.3 Image Registration Algorithm

The proposed algorithm is an hybrid system (see Fig.2), implemented in a pyramidal way.

Images A and B are the inputs of the system. Between two pyramid levels, a scaling function allows to compute the transformation of an upper level into the reference frame of a lower level. The output of the system is $\mathscr{T} = T_{L_0}$, the desired transformation between images A and B. At level L_N, the ESM-Mestimator gives an estimation of the transformation and a confidence coefficient $Conf_{ESM}$ is calculated. This coefficient is obtained by a ZNCC (Zero-mean Normalized Cross-Correlation) weighted by the M-Estimator weights W_{ESM}. About the algorithmic control, if this coefficient is not above a threshold ($Conf_{ESM} \leq \tau$), the ESM is replaced by a movement estimator using the interest corners matching. This coefficient is then estimated in function of the inliers (points that respect the global motion model) given by the RANSAC algorithm and the quality of the matching. Indeed, the use of a ZNCC is not possible in this second case because it would not be weighted and then erronated because of the outliers (points that do not respect the global motion model). If the geometric approach did not perform well, then the output is $T_{L_N} = \hat{T}_{L_N}$ (equal to input).

With this coupling, we get both the advantages of the global and geometric approaches. It gives to our method robustness against large movements, against moving targets and occlusions , high precision and a low computation time. Moreover, the auto-evaluation system(Bonnin et al.,) allows the evaluation of the estimated transformation confidence.

(Ladikos et al., 2007) also use such an hybrid system. Main differences are the fact that they use separately the two approaches in a different state of their FSM (Finite State Machine) whilst our proposed system combine the two approaches at each level of the pyramid in a same process. Moreover, they use NCC to validate the results whereas we prefer a WZNCC (Weighted ZNCC) to be robust to occlusions and noise by using the weights of the M-Estimator.

5 MOVING TARGET DETECTION

5.1 Prediction using Blob Detection

The weight image W_t, output of the image registration step, is segmented using a blob detection ((Rosenfeld and Pfaltz, 1966)). It returns potential moving targets. Information of sizes and centroids of targets are used for the verification step.

5.2 Verification with Robust Spatio-temporal Tracking

Main Objective. The use of information coming from two consecutive images is not enough for a robust tracking. It only allows to detect potential targets,

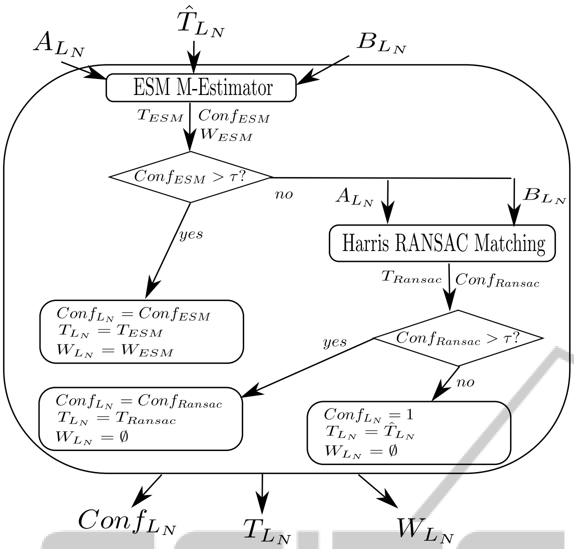


Figure 2: Image registration algorithm for a level L_N of the pyramid.

whose certain are false positives. The objective of the tracker is to reject them.

This tracker aims at checking the spatio-temporal coherency (from N images, say N=10) of positions of the moving targets of various sizes (from 3*3 pixels to 50*50 pixels for a VGA image).

When looking at target detection for a short time in a same reference frame, we noticed that centroids of moving targets leave like "footprints" with spatio-temporal coherency (see Figure 7 b)).

A Temporal Tracker... For a potential target T_p detected at time $t-N$, the algorithm can not decide if it is an actual moving target or a false positive until t, with $N = 10$ images in all of our tests. After this time of tracking, regarding the polyline trajectory of the target, it decides to label T_p as a true target T.

The decision is done by assuming a target is moving with a smooth trajectory. All potential targets which are impossible to track with the polyline are rejected.

All positions of targets and false positives are stored for the N previous images. This is the temporal sliding window. Each target or false positive of the oldest time ($t-N$) initializes several polylines whose the smoothest is retained, if it is smooth enough.

A Spatial Tracker... In order to look for the best polyline of a moving target, we have to compensate the camera motion during the temporal sliding window. This is easy to calculate thanks to the estimation between two consecutive frames given by the image registration block.

By doing this, we can filter the hypothetic detections by only keeping whose contained in a research area (represented with circles in Figure3). The smaller circle contains hypothetic detections for time $t-N+1$ ($N=5$). The larger circle contains hypothetic detections for current time t.

This step of registration in the first frame of the temporal sliding window is very important because of the large movements that camera can make.

In presence of wind, the UAV motion is erratic. This compensation allows to transform the target trajectory in a soft one.

The Figure 3 describes our method for $N = 5$. Starting from the target or false positive T_{t-5} at position $\begin{bmatrix} u_T & v_T \end{bmatrix}_{t-5}^{T}$. At $t-4$, three possibilities: $C_{1_{t-4}}$, $C_{2_{t-4}}$ and no detection. This last possibility is due to the fact that the target may be occluded during a short time. At $t-3$, only the possibilities allowing a continuity in angles are taken into account, and no detection. The process is iterated until current time t.

The polyline allowing the best compromise between real target detection and smoothing trajectory is kept. Two possible polylines are shown in figure 3. The one in continuous line is kept.

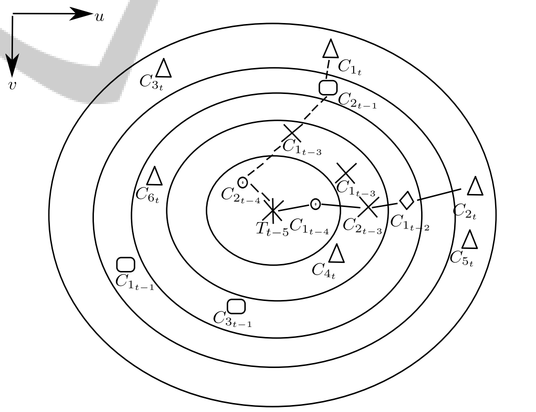


Figure 3: Spatio temporal-tracker.

6 RESULTS

6.1 Image Registration

6.1.1 M-Estimator

The M-Estimator avoids the deterioration in the precision of the estimated transformation in presence of moving targets or occlusions. To check this, we simulated a motion of the camera along the u-axis between two images. Then, we added a percentage of rectangular moving targets with random color for each object. The figure 4 shows the results : the error of the translation estimation function of the moving targets percentage, with and without M-Estimator. For small targets (only few pixels), with a very low motion (near subpixel), it is impossible to detect them with ESM only, without the using of M-Estimator weighting whereas this is possible with the ESM M-Estimator (see Figure 4).

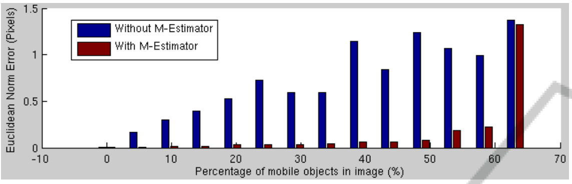


Figure 4: Estimation error function of moving pixels percentage.

6.1.2 Weight Image W_t

An example of weights image result is given in figure 5, as a heat map image. Thanks to the high precision in the estimation of the transformation, small moving targets are detected with precision.

Figure 5: Heat map image based on M-Estimator weights W_t.

Even if registration gives good results, on certain images some wrong detections still remain (see Figure 6). We can see these problems, due to compression, bad image quality, pixels interpolation (problem localized on gradients essentially)...

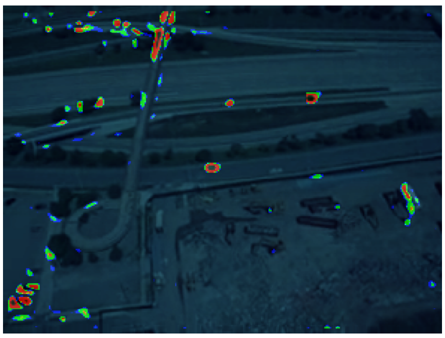
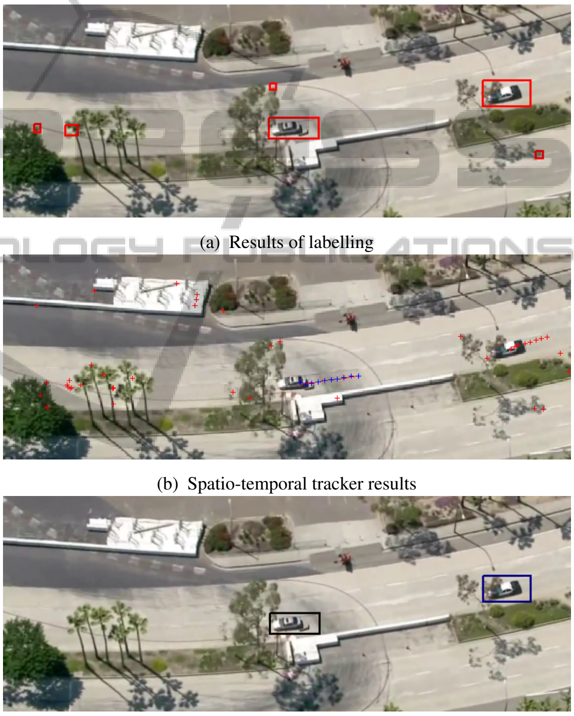


Figure 6: Some false positives in a heat map image based on M-Estimator weights W_t. To see the original image, look at Figure 8, Sequence 2.

We noticed that 40% false positives are present before filtering with the robust spatio-temporal tracker and 0% of false negatives (undetected targets). These results indicate that we need to filter to eliminate false positives detections. However, the fact that all moving targets are detected in different video sequences proves the very good results of image registration.

6.2 Spatio-temporal Tracking

This filter allows us to get only 3% of false positives as a result of GMTI. The figure 7 shows the importance of such a filter.



(a) Results of labelling

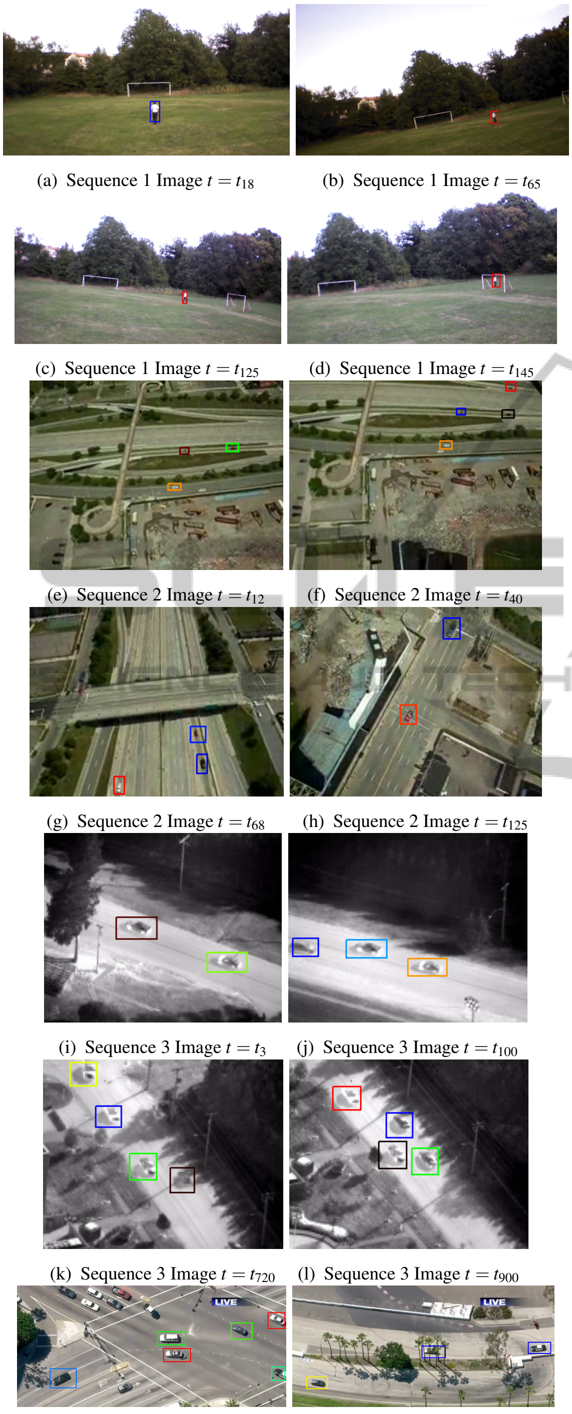
(b) Spatio-temporal tracker results

(c) Results after filtering with the tracker

Figure 7: Spatio-temporal tracker used for filtering real moving targets and wrong detections. All the possible detections are registered in a same reference frame as we can clearly view the "footprints" of all the targets (b). The correct polyline is displayed with blue crosses. Red crosses indicate the outliers for the current target tracked.

6.3 Ground Moving Target Indicator

Figure 8 shows the results of the GTMI for different kinds of videos taken by cameras mounted on UAVs. Thanks to the design of our whole process (working at 30Hz with VGA images and Proc. I5 2Ghz), even targets of 3×3 pixels size are detected as moving targets. This allows us not to miss any moving targets.



(a) Sequence 1 Image $t = t_{18}$ (b) Sequence 1 Image $t = t_{65}$

(c) Sequence 1 Image $t = t_{125}$ (d) Sequence 1 Image $t = t_{145}$

(e) Sequence 2 Image $t = t_{12}$ (f) Sequence 2 Image $t = t_{40}$

(g) Sequence 2 Image $t = t_{68}$ (h) Sequence 2 Image $t = t_{125}$

(i) Sequence 3 Image $t = t_3$ (j) Sequence 3 Image $t = t_{100}$

(k) Sequence 3 Image $t = t_{720}$ (l) Sequence 3 Image $t = t_{900}$

(m) Sequence 4 Image $t = t_3$ (n) Sequence 4 Image $t = t_{300}$

Figure 8: Results of target detection on a real images sequence.

7 CONCLUSIONS AND FUTURE WORK

We proposed a real time method for Ground Moving Target Indicator from a camera mounted on a UAV. We designed a whole process that we detailed in this article, composed of two main parts : an image registration block to compensate camera motion, and a moving target detecion block to predict and validate the position of moving targets in the current image. We tested this GTMI algorithm for different kinds of video sequences. Thanks to the design of a sub-pixellic robust registration algorithm and the spatio-temporal detection-based tracker, even targets of 3×3 pixel size with low speed are detected as moving targets.

About future work, we can notice that this paper does not focus about an accurate segmentation and labelling of the moving targets. This task is however very important because it could give us better results for the polyline tracker, and moreover allows us to display with a better precision the moving targets found.

REFERENCES

A.Goshtasby, G.C. Stockman, C. (1986). A region-based approach to digital image registration with subpixel accuracy. *IEEE Transactions on geoscience on remote sensing*, 24:390–399.

Ali, S. and Shah, M. (2005). Cocoa - tracking in aerial imagery. In *Proc. Int. Conf. on Computer Vision*.

Baker, S. and Matthews, I. (2002). Lucas-kanade 20 years on: A unifying framework: Part 1. (CMU-RI-TR-02-16).

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359.

Benhimane, S. and Malis, E. (2004). Real-time image-based tracking of planes using efficient second-order minimization. In *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 943–948.

Bonnin, P., Blazevic, P., Morillon, J., Fialaire, C., and Benoist, J.-S. Real time tracking system by vision simplifying the interaction between human and robot for remote control of mobile robot. In *RO-MAN*, pages 945–950. IEEE.

chih Wang, C., Thorpe, C., Hebert, M., Thrun, S., and Durrant-whyte, H. (2004). Simultaneous localization, mapping and moving object tracking.

Cohen, I. and Medioni, G. (1998). Detection and tracking of objects in airborne video imagery.

Crivelli, T., Bouthemy, P., Cernuschi-Frias, B., and Yao, J.-F. (2011). Simultaneous motion detection and background reconstruction with a conditional mixed-state markov random field. *International Journal of Computer Vision*, 94(3):295–316.

Dame, A. and Marchand, E. (2012). Second order optimization of mutual information for real-time image registration. *IEEE Trans. on Image Processing*, 21(9):4190–4203.

de Cabrol, A., Bonnin, P. J., Hugel, V., Blazevic, P., and Chetto, M. (2005). Video rate color region segmentation for mobile robotic applications. *Proc. SPIE*, 5909:59091E–59091E–12.

de Cabrol, A., Bonnin, P. J., Hugel, V., Bouchefra, K., and Blazevic, P. (2005). Temporally optimized edge segmentation for mobile robotics applications. In *Applications of Digital Image Processing XXVIII*, volume 5909 of *SPIE*, pages 448–459.

Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395.

Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *In Proc. of Fourth Alvey Vision Conference*, pages 147–151.

Hartley, R. I. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition.

Huber, P., Wiley, J., and InterScience, W. (1981). *Robust statistics*. Wiley New York.

Irani, M. and Anandan, P. (1998). A unified approach to moving object detection in 2d and 3d scenes. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 20.

Klose, S., Heise, P., and Knoll, A. (2013). Efficient compositional approaches for real-time robust direct visual odometry from rgb-d data. In *IROS*, pages 1100–1106. IEEE.

K.Poornima, R. (2012). A method to align images using iamge segmentation. *IJCSE*, 2.

Ladikos, A., Benhimane, S., and Navab, N. (2007). A realtime tracking system combining template-based and feature-based approaches. In *IN VISAPP*.

Lin, Y., Yu, Q., and Medioni, G. (2011). Efficient detection and tracking of moving objects in geo-coordinates. *Mach. Vision Appl.*, 22(3):505–520.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110.

Malis, E. and Marchand, E. (2006). Experiments with robust estimation techniques in real-time robot vision. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'06*, pages 223–228, Beijing, Chine.

Migliore, D., Rigamonti, R., Marzorati, D., Matteucci, M., and Sorrenti, D. G. (2009). Use a single camera for simultaneous localization and mapping with mobile object tracking in dynamic environments.

Odobez, J. and Bouthemy, P. (1997). Separation of moving regions from background in an image sequence acquired with a mobile camera. In Li, H., Sun, S., and Derin, H., editors, *Video Data Compression for Multimedia Computing*, chapter 8, pages 283–311. Kluwer Academic Publisher.

Reilly, V., Idrees, H., and Shah, M. (2010). Detection and tracking of large number of targets in wide area surveillance. In *Proceedings of the 11th European Conference on Computer Vision Conference on Computer Vision: Part III*, ECCV'10, pages 186–199, Berlin, Heidelberg. Springer-Verlag.

Richard Evans, E. T. (2007). Visual mti for uav systems. *4th EMRS DTC Technical Conference Edinburgh*.

Rodrguez-Canosa, G. R., Thomas, S., del Cerro, J., Barrientos, A., and MacDonald, B. (2012). A real-time method to detect and track moving objects (datmo) from unmanned aerial vehicles (uavs) using a single camera. *Remote Sensing*, 4(4):1090–1111.

Rosenfeld, A. and Pfaltz, J. L. (1966). Sequential operations in digital picture processing. *J. ACM*, 13(4):471–494.

Smith, S. M. and Brady, J. M. (1995). Susan - a new approach to low level image processing. *International Journal of Computer Vision*, 23:45–78.

yves Bouguet, J. (2000). Pyramidal implementation of the lucas kanade feature tracker. *Intel Corporation, Microprocessor Research Labs*.

Zitov, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000.