XVI Reunión de Trabajo en Procesamiento de la Información y Control, 6 al 9 de octubre de 2015

Use of Inertial and Altimeter Information for Rectified Searches in Image Target Tracking for Drone Applications

María Curetti †, Mariano Serra †, Santiago Garcia Bravo †, Ana Georgina Flesia •, Ladislao Mathé ‡

† Centro de Investigaciones Aplicadas, DIGID, Fuerza Aérea Avenida Fuerza Aérea 6500, Córdoba, Argentina.
curetti.maria@gmail.com
marianoserra88@gmail.com
sgbravo@iua.edu.ar
• Centro de Investigación y Estudios en Matemática de Córdoba, Consejo Nacional de Investigaciones Científicas y Técnicas; Universidad Nacional de Córdoba, Facultad de Matemática, Astronomía y Física
gflesia@gmail.com
‡ Universidad Nacional de Córdoba, Facultad de Ciencias Exactas, Físicas y Naturales
ing.ladislao.mathe@gmail.com

***Abstract*— In this paper, a model for target tracking with data fusion taken from inertial and altimeter sensors for a quadcopter is proposed. With some environmental restrains and using INS/altitude data, the new position and template changes are predicted. The model dynamics is estimated on the basis of the specifications of the UAV IRIS+, with a pointgrey blackfly camera (PGE-13E4C-CS), a gimbal (TAROT 2D), and a Laser Rangefinder (SF02).Simulation results are shown.**

***Key words*— Image processing, Sensor fusion, Object tracking, and Image wrapping.**

1. INTRODUCTION

Object tracking through image sequences is one of the problems of image processing most often studied. Besides its inherent worth, tracking is the first step for any navigation system based on images. A wide variety of algorithms is used to model the object of interest and to then look for it in the image. One of the first methods to be applied in object tracking is template matching. This algorithm takes a known intensity mask of the item being tracked and compares it with the image until it finds the place where their correlation takes its maximum value [1], [2], [3], [5].

This method presents a series of problems, being the change in the object appearance over time one of the most significant. Therefore, variations in the light and in the shape of the image must be taken into account. To address this, a variety of approaches with different amount of computational load have been employed. [4], [7]

Updating the template after a certain number of frames is a basic method often applied to outline a valid solution for some of these problems. The main failure of this algorithm is that the object may drift from the mask. To keep the element centered in the template, the approach proposed by Matthews et al [6] is to preserve the first mask and to use it in the new template to prevent the drift. Transitory occlusion may occur deteriorating the mask in algorithms that lack robustness. Nguyen et al [8], suggest the use of Kalman filters to estimate the intensity of the new template so as to improve robustness against occlusions. In newer approaches, for example in [10], matching based on local brightness is replaced by a search for a model, thus avoiding failure caused by light changes.

At the moment, Kanade-Lucas-Tomasi (KLT) is one of the most prevalent tracking methods that use template image alignment techniques. It has been extensively studied in the seminal work of Lucas and Kanade (1981), Shi and Tomasi (1994), and in the unifying framework of KLT variants by Baker and Matthews (2004).

Most template matching algorithms for target tracking are based on the basic assumption that target appearance changes slowly over time. Consequently, they are vulnerable to fast rotation or severe camera shaking.

This type of disturbances is frequent in UAV vision applications. Tracking algorithms, in general, use a template warping model for template matching along successive frames in a video sequence [2, 7, 25]. The choice of warping model plays a significant role in the performance of the tracker. The homography-based warping model is a parametrization with 8 independent variables. However, when the camera is exposed to large movements the true position and rotation may be too different from the initial value acquired by the camera on the basis of the previous frame, and as a consequence it may be difficult for the algorithm to converge. Therefore, an adequate estimate of the initial parameters is critical for coping with severe image deformation.

The advantage in this kind of applications is that information of inter frame attitude and position change is available before the new frame appears. Almost all UAV systems are equipped with inertial systems, altimeters, GPS and motion estimators. So once these sensors are calibrated, the initial parameters can be estimated more accurately, and a more robust algorithm can be achieved.

In this work, a tracking method that uses the template matching algorithm developed in [13] is proposed. This enhanced algorithm includes inertial and altitude sensor data fusion to improve the initial estimate of the target position and attitude relative to the camera. The algorithm developed is to be implemented in the near future on aNvidia Jetson TK1 which will control a system made up of an IRIS+ UAV, with a pointgrey blackfly camera (PGE-13E4C-CS), a gimbal TAROT 2D, and a SF02 Laser Rangefinder, (see Fig. 1).The simulation results will be presented showing the robustness of the enhanced method which uses an algorithm which considers image data alone. The dataset introduced for the simulations is the one used in the work by Warren et al [33].

Figure 1: Full system prototype

2. DESCRIPTION OF THE WORK

2.1. Background

The use of inertial sensors in machine vision applications was proposed more than twenty years ago. Further studies have investigated the cooperation between inertial and visual systems in autonomous navigation of mobile robots, in image correction and as a way to improve motion estimation for 3D reconstruction of structures. More recently, a framework for cooperation between vision sensors and inertial sensors has been proposed. The use of gravity as a vertical reference allows the calibration of focal length camera with a single vanishing point, the segmentation of the vertical and the horizontal plane. In the work by J. Lobo and J. Dias [14] a calibration method is presented which uses a function for detecting the vertical reference (gravity) and 3D mapping, and in [15], such vertical inertial reference is used to improve the alignment and registration of the depth map. Applications of this method in robotics are increasing. Initial works on vision systems for automated passenger vehicles have also incorporated inertial sensors and have explored the benefits of visual-inertial tracking [16], [17]. Other applications include agricultural vehicles [18], wheelchairs [19] and robots in indoor environments [20], [21]. Other recent works include an application of this method in Unmanned Aerial Vehicles (UAV) [22], [23].

In inertial-aided visual tracking, many approaches have been studied. Kaushik [24] showed a scheme which compensates for perturbations by using 3-axis inertial sensors, and Kanade [25] developed a novel inertial-aided Kanade-Lucas-Tomasi (KLT) feature tracking method using gyroscope data. In most cases, angular motion compensations are performed using inertial sensors. To compensate the translational motion, accelerometers data is used [26]. Recently, the Vestibulo-Ocular Reflex (VOR) based vision stabilization systems show fast and accurate stabilization performances [27]. Cho et al. presented the VOR-based target tracking system using accelerometer information [28]. Their paper computes the translational motion of the robot by using accelerometer information. On the basis of this information, a vision sensor mounted on the robot rotates towards the selected target, periodically compensating errors from visual information. (High Performance Vision Tracking system)

In [29], a 8-DOF affine photometric model has been applied. Due to its high computational complexity, a restrained Hessian update in KLT as well as GPU (graphical processing unit) acceleration are used, achieving 1024 feature track at video rate in a NVIDIA GeForce 8800 GTX.

Tanathong and Lee, [32] propose two methods. One that rotates the new frame taking into account the inter frame rotation data, and another one that rotates only the window where the target position is predicted. These two methods are used to enhance the KLT affine model. The proposed methods proved to be more efficient and accurate than the KLT affine option. This reason together with its low computational burden have led to the selection of the second algorithm for the enhancement of the TV tracker of Curetti et al. It is important to take into account that the system proposed must be a simple system so that it can be implemented on the Nvidia Jetson TK1 at 60 fps.

2.2. Previous work

This section describes the tracking algorithm proposed by Curetti et al. [13] and the second method proposed by Tanathong and Lee, [32] to enhance the KLT affine search.

2.2.1. Adjustable Tracking Algorithm with Adaptive Template Matching

The tracking algorithm proposed [13] consists of four sequential steps. First, the best match is found taking into account a normalized template correlation and a position predictor. Then, a new template size and center is estimated using the image edge map. Next, the result of that last step is filtered. Finally, a new template is calculated for the next iteration.

In simple terms, the normalized template is the target template without its mean value:

$$t'(x,y) = t(x,y) - \bar{t} \tag{1}$$

where t' is the normalizer template, t is the target template and $\bar{t}$ is the template mean.

The template update is simply an IIR filter applied to the target mask obtained in the image sequence and aligned with the normalized correlation.

2.2.2. Translation-Based KLT Tracker with Rotated Tracking Window

This solution is based on the underlying concept of the KLT algorithm with an affine model. Since its focus is on dealing with the camera rotation, it only considers the rotation factor while other affine parameters are discarded. Thus, instead of estimating all affine warping parameters, as it occurs in the original affine-based tracker, the initial values of the rotation matrix are kept constant and only the translation vector is determined. Based on this assumption, the warping function is expressed as equation (2). This motion model reduces the computational burden of solving six affine parameters to estimate only two translation parameters, denoted as p, where p is defined as ⟦(dx,dy)⟧^T

$$W(x,p) = \begin{pmatrix} \cos\theta & \sin\theta & d_x \\ -\sin\theta & \cos\theta & d_y \end{pmatrix} (x \quad y \quad 1)^T. \tag{2}$$

This proposed solution may be viewed as if a translation-based KLT algorithm were performed using the tracking window that has been rotated for a constant θ degree.

2.3. Proposed tracking system

In this work, the homographic transformation between two consecutive frames taken from a UAV facing the ground is estimated. The position and attitude of the camera in each frame is approximated using inertial sensors data and the ground is modeled as a plane. With this homographic transformation, the tracking window or Region of Intererst (ROI) from the Adjustable Tracking Algorithm proposed in [13] is mapped to the new image (Fig. 2) and the mask is adjusted with the rotation angle. The rest of the tracking algorithm is performed as explained in 2.2.1. The only modification relative to Tanathong and Lee´s algorithm is that the mask size is kept constant.

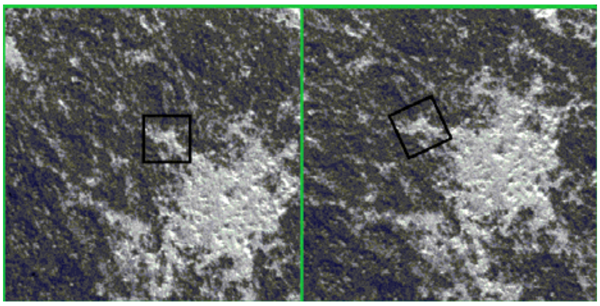

Figure 2: ROI projection from one frame to the next.

2.4. Simulation results

Through the use of simulations, the efficacy of the proposed method is compared to the one developed by Curetti et al. in [13]. The data used to simulate both tracking systems is found in https://wiki.qut.edu.au/display/cyphy/Kagaru+Airborne+Dataset.

They were obtained with the following experimental setup:

- a 1/3 scale Piper Cub with a 3.6-m wing span of and a 2.3m-fuselage length , capable of attaining speeds of 30 to 110km/h with a maximum payload of 6kg.
- an off-the-shelf mini-ITX computer system running an Intel Atom processor (1.6GHz), with two 64GB solid-state drives in a RAID0 configuration
- an IEEE1394B color Point Grey Flea 2 camera. The camera is placed in the fuselage of the platform, facing downwards towards the terrain,
- a 6-mm lens is used with a field of view of approximately 42°×32°

Data was collected over a 90 second portion of flight, at an altitude of 20-100m and a speed of 20m/s. Bayer encoded color images were logged at a resolution of 1280×960 pixels at 30Hz. Shutter time for each frame was set at 8.5µS to counteract motion blur. The area was rural farmland with relatively few trees, animals and buildings.

An XSens MTi-G INS/GPS system was used as the INS/GPS measurement system. GPS, unfiltered IMU data and filtered INS pose were recorded at 120Hz from the XSens MTi-G.

To make the simulation data more similar to the dynamic of the application proposed (rotorcraft UAV), the images were transformed simulating faster trajectories. The main change observed for different trajectories was the yaw maximum angular speed. As the maximum speed of the rotorcrafts UAV is around 200º/seg, the following velocities were simulated: 3, 10, 17, 24, 31, 45 and 59º/seg.

The targets used to evaluate both systems were the same (20 random targets). The results are shown in Figures 1, 2, 3 and 4.

In Fig. 3 the efficacy of the proposed system (corrected with the inertial data) is compared with the original method proposed by Lobo and Dias [14] (uncorrected). We define efficacy as the number of targets that were followed correctly during the 1.3 second of simulated trajectory over the total of targets in the sample. The proposed method is more robust to violent movements and has an efficacy of more than 90%, even at a yaw angular speed of 59º/seg. The uncorrected method efficacy is high when the movements are slow; however, for faster movements, its efficacy falls.

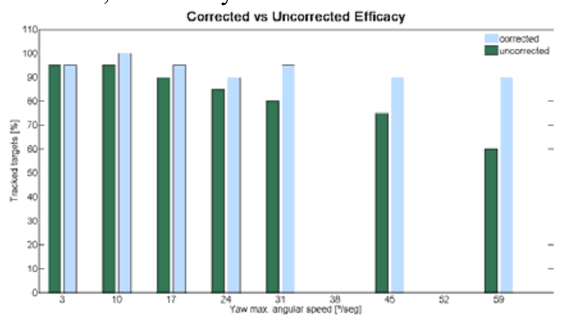

Figure 3 Comparison between the efficacy of the proposed system (corrected) and the original method (uncorrected) for different yaw maximum speed

In Fig. 4, the horizontal component of the tracked trajectory of the 20 targets is shown for the uncorrected system in a trajectory with a maximum yaw angular speed of 3º/seg. From the sample of 20 targets, only one cannot be followed (notice the sudden change in one of the trajectories).

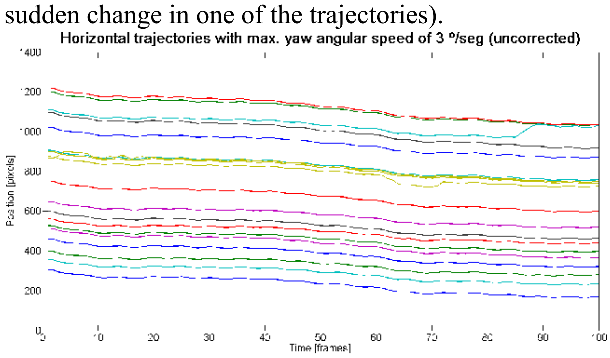

Figure 4 Horizontal component of the tracked trajectory of the 20 targets for the uncorrected system, with a maximum yaw angular speed of 3º/seg

In Fig. 5, the vertical component of the tracked trajectory of the 20 targets is shown for the uncorrected system in a trajectory with a maximum yaw angular speed of 59º/seg. From the sample of 20 targets, 8 targets cannot be followed.

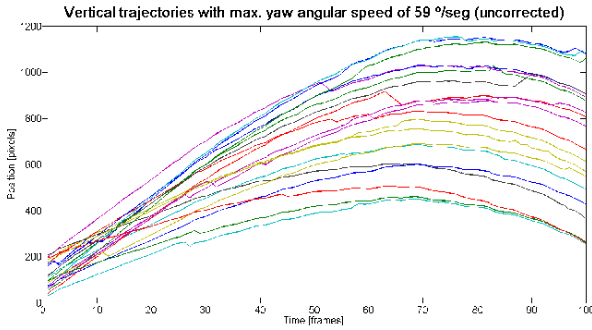

Figure 5 Vertical component of the tracked trajectory of the 20 targets for the uncorrected system, with a maximum yaw angular speed of 59º/seg

In Fig. 6, the vertical component of the tracked trajectory of the 20 targets is shown for the corrected system in a trajectory with a maximum yaw angular speed of 59º/seg. From the sample of 20 targets, only two cannot be followed

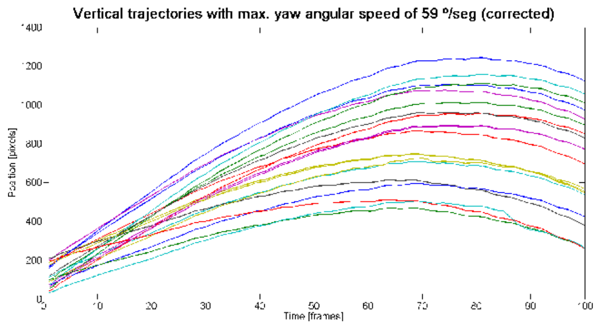

Figure 6 Vertical component of the tracked trajectory of the 20 targets for the corrected system, with a maximum yaw angular speed of 59º/seg

3. CONCLUSIONS AND FUTURE DIRECTIONS

The proposed method is more robust to violent movements and has an efficacy of more than 90%, even at a yaw angular speed of 59º/seg. The uncorrected method can be used if the dynamics of the application are slow (less than 24º/seg. of yaw angular speed), However, its efficacy falls when the movement is faster. The accuracy of the systems depends on the application, but more specifications are needed if a method is to be chosen. In the future, the proposed method will be evaluated using data collected with this lab experimental platform. Also, the tracking system will be implemented with an on board image processing on the Nvidia Jetson TK1.

REFERENCES

[1] Black, M., Jepson, A. "Eigen-Tracking: Robust Matching and tracking of Articulated Objects Using a View-Based Representation". *Int'l. J. Computer Vision 36*: pp. 63-84, 1998.

[2] Hager, G., Belhumeur, P. "Efficient Region Tracking with Parametric Models of Geometry and Illumination". *IEEE Trans. Pattern*

Analysis and Machine Intelligence 20: pp. 1025-1039, 1998.

[3] Cootes, T., Edwards, G., Taylor, C. "Active Appearance Models". *IEEE Trans. Pattern Analysis and Machine Intelligence 23*: pp. 681-685, 2001.

[4] Xiao, J., Kanade, T., Cohn, J.: "Robust Full-Motion Recovery of Head by Dynamic Templates and Re-Registration Techniques". *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 163–169 (2002).

[5] Baker, S., Matthews, I. "Lucas-Kanade 20 Years on: A Unifying Framework". *Int'l. J. Computer Vision,* 53: pp. 221-255, 2004.

[6] Matthews, I., Ishikawa, T., Baker, S.: "The Template Update Problem". *IEEE Trans. on Pattern Analysis and Machine Intelligence* 26(6), 2004.

[7] Zhong, Y., Jain, A.K., Dubuisson-Jolly, M.-P.: "Object Tracking Using Deformable Templates". *IEEE Trans. on Pattern Analysis and Machine Intelligence* 22(5), 2000.

[8] Nguyen, H.T., Worring, M., van den Boomgaard, R.: "Occlusion robust adaptive template tracking". *Eighth IEEE International Conference on Computer Vision,* 2001.

[9] Comaniciu, D., Ramesh, V., Meer, P.: "Real-Time Tracking of Non-Rigid Objects using Mean Shift". *IEEE,* 2000

[10] Wu, Y., Fan, J.: "Contextual Flow". *IEEE Conference on Computer Vision and Pattern Recognition,* 2009

[11] Lucas BD and Kanade T "An iterative image registration technique with an application to stereo vision". pp. 674–679 *Proceedings of the 7th International Joint Conference on Artificial Intelligence, Vancouver, Canada*,. 1981

[12] Shi J and Tomasi C "Good features to track". "*IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94), Seattle, WA.* 1994

[13] Curetti, M., Garcia Bravo, S., Arri, G. S., & Mathe, L. "Adjustable Tracking Algorithm with Adaptive Template Matching". pp34-39, *Latin America Transactions, 11*(1),2013.

[14] J. Lobo and J. Dias, "Vision and inertial sensor cooperation using gravity as a vertical reference," pp. 1597–1608. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 25, no. 12*,, Dec. 2003.

[15] Lobo, J., & Dias, J. "Inertial Sensed Ego-motion for 3D Vision". *Journal of Robotic Systems*, 21(1), 3-12. 2004

[16] E. D. Dickmanns, "Vehicles capable of dynamic vision: a new breed of technical beings" *Artif. Intell., vol. 103, no. 1-2*, pp. 49–76, Aug. 1998.

[17] J. Goldbeck, B. Huertgen, S. Ernst, and L. Kelch, "Lane following combining vision and dgps," *Image and Vision Computing, vol. 18, no. 5, pp. 425–433*, 2000.

[18] T. Hague, J. A. Marchant, and N. D. Tillett, "Ground based sensing systems for autonomous agricultural vehicles," *Computers and Electronics in Agriculture, vol. 25, no. 1-2, pp. 11*–28, 2000.

[19] T. Goedem, M. Nuttin, T. Tuytelaars, and L. V. Gool, "Vision based intelligent wheel chair control: The role of vision and inertial sensing in topological navigation," *Journal of Robotic Systems*, vol. 21, no. 2, pp. 85–94, 2004.

[20] D. D. Diel, P. DeBitetto, and S. Teller, "Epipolar constraints for visionaided inertial navigation," *Proc. IEEE Motion and Video Computing, pp. 221–228*, 2005.

[21] I. Stratmann and E. Solda, "Omnidirectional vision and inertial clues for robot navigation," *Journal of Robotic Systems, vol. 21, no. 1, pp. 33–39*, 2004.

[22] L. Muratet, S. Doncieux, Y. Briere, and J.-A. Meyer, "A contribution to vision-based autonomous helicopter flight in urban environments," *Robotics and Autonomous Systems, vol. 50, no. 4, pp. 195–209*, 2005.

[23] P. Corke, "An inertial and visual sensing system for a small autonomous helicopter," *Journal Field Robotics, vol. 21, no. 2, pp. 43–51*, 2004.

[24] R. Kaushik, M. Marcinkiewicz, J. Xiao, S. Parsons, and T. Raphan, "Implementation of bio-Inspired Vestibulo-Ocular Reflex in a quadrupedal Robot," *Proc. IEEE Int. Conf. Robotics and Automation, pp 4861-4866*, 2007.

[25] M. Hwangbo, J. –S. Kim, and T. Kanade, "Inertial-aided KLT feature tracking for a moving camera," *Proc. IEEE/RJS Int. Conf. Intelligent Robots and Systems, pp. 1909-1916,* 2009

[26] Z. Jia, A. Balasuriya, and S. Challa, "Sensor fusion based 3d target visual tracking for autonomous vehicles with the out-of sequence measurements solution," *Proc. IEEE Int. Conf. Robotics and Automation, pp. 1841-1846,* 2005.

[27] A. Lenz, T. Balakrishnan, A. G. Pipe, and C. Melhuish, "An adaptive gaze stabilization controller inspired by the vestibulo-ocular

reflex," *Bioinspiration & Biomimetics, vol.3, pp. 1-11*, 2008.

[28] H. K. Ouh, W. Bahn, J. Park, W. S. Hwang, H. I. Kwon, K. S. Kim, and D. I. Cho, "VOR based target tracking system with an accelerometer for mobile robots," *Proc. 6th Int. Conf. Ubiquitous Robots and Ambient Intelligence, pp. 133-137*, 2009.

[29] Hwangbo, M., Kim, J. S., & Kanade, T.. "Gyro-aided feature tracking for a moving camera: fusion, auto-calibration and GPU implementation". *The International Journal of Robotics Research*, 0278364911416391. 2011

[30] Park, J., Hwang, W., Kwon, H. I., Kim, J. H., Lee, C. H., Anjum, M. L., & Cho, D. I. "High performance vision tracking system for mobile robot using sensor data fusion with Kalman filter". *Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on* (pp. 3778-3783). *2010.*

[31] Vincenzo Angelino, C., Baraniello, V. R., & Cicala, L. (2013, July). "High altitude UAV navigation using IMU, GPS and camera". *Information Fusion (FUSION), 16th International Conference on* (pp. 647-654). IEEE. *2013*

[32] Tanathong, S., & Lee, I. "Translation-Based KLT Tracker Under Severe Camera Rotation Using GPS/INS Data". *Geoscience and Remote Sensing Letters, IEEE*, *11*(1), 64-6, 2014.

[33] Warren, M., McKinnon, D., He, H., Glover, A., Shiel, M., & Upcroft, B. "Large Scale Monocular Vision-only Mapping from a Fixed-Wings UAVs". *Field and Service Robotics* (pp. 495-509). Springer Berlin Heidelberg, 2014.