

UAV VISUAL AUTOLOCALIZATION BASED ON AUTOMATIC LANDMARK RECOGNITION

P. Silva Filho^a, E. H. Shiguemori^a, O. Saotome^b

^aIEAV - Instituto de Estudos Avançados, São José dos Campos-SP, Brazil - (silvafilho, elcio)@ieav.cta.br

^bITA - Instituto Tecnológico de Aeronáutica, São José dos Campos-SP, Brazil - osaotome@ita.br

KEY WORDS: ORB, AKAZE, Position, Estimation, GNSS, GPS-denied, Visual navigation

ABSTRACT:

Deploying an autonomous unmanned aerial vehicle in GPS-denied areas is a highly discussed problem in the scientific community. There are several approaches being developed, but the main strategies yet considered are computer vision based navigation systems. This work presents a new real-time computer-vision position estimator for UAV navigation. The estimator uses images captured during flight to recognize specific, well-known, landmarks in order to estimate the latitude and longitude of the aircraft. The method was tested in a simulated environment, using a dataset of real aerial images obtained in previous flights, with synchronized images, GPS and IMU data. The estimated position in each landmark recognition was compatible with the GPS data, stating that the developed method can be used as an alternative navigation system.

1. INTRODUCTION

Unmanned Aerial Vehicles (UAV) is one of the main strategic technologies nowadays due to their high applicability in several areas, such as urban areas and frontiers surveillance (Blumenau et al., 2013); object and landmarks recognition (Zhao et al., 2013); and remote sensing (Pajares, 2015). This popularity grew mostly because of their ability to deploy a mission with little human interaction, in other words, their applicability as autonomous systems.

The first aspect for the development of an autonomous system is regarding its navigation control. The navigation consists in obtaining information regarding the flight, the field and the aircraft itself, in order to reach a specific location (Dumble and Gibbens, 2014). Most of the autonomous navigation systems available nowadays use the Inertial Measurement Unit (IMU) and a Global Navigation Satellite System (GNSS), such as the Global Position System (GPS). Even though for most environments the GNSS+IMU navigation works well in clear sky-view, it is not a fully reliable system. Most IMU used in UAV lose precision in a short period of time, so it needs another position estimator, in order to work properly again. The GNSS is an external signal and it can be lost for many reasons such as: satellite disposition; multipath; changes in the Ionosphere and the presence of ionospheric bubbles (Muella, 2008, Takahashi et al., 2015); and jamming or signal blocking (LeMieux, 2012). In such a system, if the GNSS data is lost, the UAV will not work properly and may even cause an accident.

Several works have been developed in order to obtain an alternative or redundant navigation system for the autonomous navigation that could deal with fails in GNSS or in GNSS-denied areas (Rady et al., 2011, Singh and Sujit, 2016). One strong candidate, which is the subject of this work, is a computer vision system that can estimate the aircraft's latitude and longitude from images captured and processed onboard during flight.

1.1 Visual Navigation Systems

Nowadays, there are several researches on Visual Navigation systems. Most studies work on ways to develop methods that can

adapt to the multiple circumstances an UAV may face during flight and can affect the visual data obtained: different weather condition; different sensors; time of day; areas over which the aircraft is located; and others. The main strategies applied to the navigation in development nowadays are Visual Odometry (Quist and Beard, 2016), Simultaneous Localization and Mapping (SLAM) (Azizi et al., 2016), Template Matching (Braga et al., 2015) and Landmark Recognition (DeAngelo and Horn, 2016). Each of them has advantages and drawbacks that must be taken into consideration during development. For example, both Template Matching and Landmark Recognition are more computationally complex and need to know the region of flight a priori, but on the other hand, works implementing them have shown higher precision than works using visual odometry or SLAM. Visual Odometry can be used in unknown areas, but, the same way as an IMU, it accumulates errors throughout the flight. SLAM is one of the most discussed and studied strategy nowadays, but it always needs to fly over previously visited areas, in order to have a precise navigation.

There are few works on landmark recognition for aerial navigation because it is a complex system and most works on this area are in a more heuristic and less practical approach (Silva, 2015). The landmark recognition system aims at finding specific chosen structures in the aerial images that are captured during the UAV flight by an onboard camera. After the recognition, the UAV location is then estimated in real-time, in order to support the navigation system to accomplish a planned mission (DeAngelo and Horn, 2016). Landmarks can be understood as salient, usually man-made infrastructures that stands out on the field, for example roads with intersections and crossings, rivers with crossing roads, runway and taxiway structures, shores of lakes, islands, large buildings, towers, bridges, and wood edges or isolated pieces of woodland, clearings in woods (Michaelsen and Meadow, 2014). Their location is well known and they are selected during mission planning, therefore, the Landmark Recognition Navigation System needs to previously know the area of the flight (Silva Filho et al., 2014). This work, then, develops a practical and direct landmark recognition system for aerial navigation, combining real-time with a high precision recognition system. The

proposed method has been tested in two different experiments: geo-referencing a videoframe using landmark recognition; and estimating an UAV position, in a simulated flight using previously obtained aerial images.

1.2 Related Works

Even though landmark recognition is not a new subject on the literature for autonomous ground vehicles (Farag and Abdel-Hakim, 2004), the approach for aerial navigation systems is not well explored yet, mostly because of its complexity and high real-time requirements on precision and computer processing (Silva Filho, 2016). Because of that high computational costs, there are some works that state the need to send the data to a ground station, which will process the images, and then send to the UAV only the result of the recognition (Silva, 2015, Michaelsen and Meadow, 2014).

There are several challenges in the recognition process, such as difference in resolution, rotating, translating, scale, luminosity, different sensors and many others. Most works on landmark recognition for aerial navigation takes on the results of already developed object-recognition algorithms and adapt them to the different aerial circumstances. In (Cruz, 2014), Histogram of Gradient (HOG) associated with a Support Vector Machine (SVM), Haar-like feature cascade and Local Binary Pattern (LBP) cascade are all applied to recognize specific classes of objects, such as soccer fields and airports. It is a suitable technique to recognize classes, but the training of the classes is time consuming. In (Kezheng et al., 2015), the UAV was able to localize an specific artificial landmark (the letter H) for navigation, detecting features and corners of the image using the Hough Transform and Labourasse Transform. The project can work on a real-time base, but is limited to recognize a specific artificial landmark. And on (Zhu and Deng, 2016), the landmarks are recognized and a mathematical scheme is proposed for distance and position estimation for the aircraft. It is still a heuristic proposal without a practical recognition algorithm.

Feature based algorithms, such as Scale Invariant Feature Transform (SIFT) (Lowe, 2004), Oriented FAST and Rotated BRIEF (ORB) (Rublee et al., 2011), AKAZE (Alcantarilla et al., 2011), have changed the object recognition field of study (Li et al., 2015). In (Lee et al., 2010) the method first extracts feature points from the image data taken by a monocular camera using the SIFT algorithm. The system selects landmark feature points that have distinct descriptor vectors among the feature points, calculate those points location and store them in a database. Based on the landmark information, the current position of the UAV is estimated. It considers as a landmark just the exact feature point instead of an object. This method has been used for indoor applications, which is a controlled environment. In outdoor flights, though, this application could not be used properly because the amount of similar features would result in a high rate of false positive encounters.

2. LANDMARK AND UAV AUTOLOCALIZATION

Finding a suitable and general method for a Landmark Recognition Navigation System is not a trivial task, mostly because of the recognition aspect. The different conditions in which the landmark may appear for the system (luminosity, rotation, scale, perspective, etc) demands a general invariant method that is quite difficult to obtain. In addition to that, there is also the requirement that the algorithm processes in real-time.

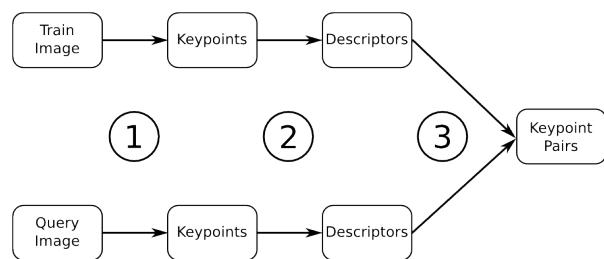


Figure 1. Feature points detection, descriptors extraction and feature matching

This work then, developed a method based on the combination of two well known object recognition methods (Feature Points and Template Matching) in order to obtain an algorithm with a high reliability that can be processed in real-time. In other words, that can provide the UAV position in such a short time, that it will still be a valid position for the autonomous navigation system. After the recognition, it was possible to extract the position information of the aircraft using the position information of the recognized landmark.

2.1 Recognition

Object recognition is a classic visual computer problem, and there are several methods developed. As it needs to be a real-time application, feature (keypoints) points detection and descriptors extraction was selected as the first most suitable strategy. There are several different feature point algorithms in the literature. Some of them were tested before deciding which would work better for the application, and the tests evaluated mostly their reliability on recognition of the landmarks but also their execution time. ORB, AKAZE, SIFT and SURF were tested. The basic structure of those algorithms is the feature point detection in the scale space of the image and the descriptor extraction of each feature point detected. This structure is performed in the image of the landmark that needs to be recognized (this one is called the train image) and in the aerial image captured during flight (the query image). The difference between each feature point algorithm is mostly on the scale-space used (linear or non-linear) and the descriptor (binary or non-binary). From the tests, the feature point with better results was AKAZE.

The feature points of each image are then matched based on their descriptors. This matching is then analyzed and graded based on a distance function. The pairs with the lower grades are considered the best matches and they are used to estimate the parameters of a General Affine Transform that maps the query image in the train image. A Fuzzy Inference System is then used to validate the Affine Transform obtained. Figure 1 shows the matching flow.

At first, only the feature point strategy was used to recognize the landmarks. During tests, it was observed a high number of false positive recognitions, which were not reduced using RANSAC or any other algorithm to reduce the outliers. RANSAC actually just added a higher processing time, with little gain in the reliability of the algorithm.

A new strategy then was developed. The result from the feature point recognition was interpreted as a candidate for the landmark. The resulted Affine Transform was the used to modify the query image and crop it to have an image of the candidate with the

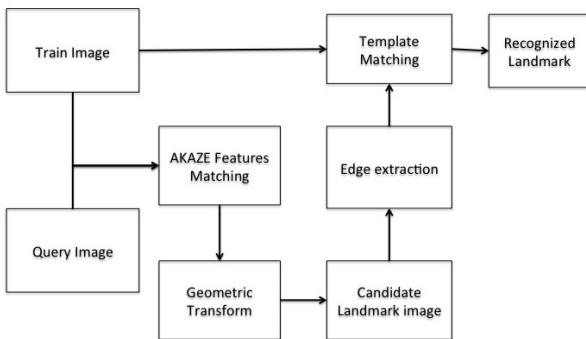


Figure 2. Proposed landmark recognition scheme

same size as the train image. If it were a true positive recognition, both images would be the same and a false positive encounter would produce a different image. In order to evaluate it, a template matching method was used. So the edges of both the candidate and the train image were extracted, resulting in two binary images. Those images would be used to compute the classic correlation coefficient using equation 1, where T is the train edge image and I is the candidate edge image. In case the correlation coefficient is higher than a threshold, the landmark is finally recognized. Figure 2 show each step for recognition.

$$R(x, y) = \frac{\sum_{x', y'} T'(x', y') I'(x', y')}{\sqrt{(\sum_{x', y'} T'(x', y')^2)(\sum_{x', y'} I'(x + x', y + y')^2)}}$$

$$T'(x', y') = T(x', y') - \frac{1}{hw} \sum_{x'', y''} T(x'', y'')$$

$$I'(x + x', y + y') = I(x + x', y + y') - \frac{1}{hw} \sum_{x'', y''} I(x + x'', y + y'')$$
(1)

2.2 Auto-localization

From the recognized landmark, it is possible to extract the latitude and longitude information for the autonomous navigation system. This position information derives from the general affine transform obtained in the recognition process. It is generated by the matched keypoints. Considering $F : T(x, y) \mapsto G(lat, long)$ the Geo-referencing relation from the pattern image T with the Object Space G , and $K : T(x, y) \mapsto Q(X, Y)$ the Geometric Transformation that maps the pattern image T in the query image Q , it is possible to build the Geo-referencing transformation H , from the query image Q , in which:

$$H : Q(X, Y) \xrightarrow{K^{-1}} T(x, y) \xrightarrow{F} G(lat, long) \quad (2)$$

It is not necessary for the pattern image to be fully geo-referenced, to do the image-to-image Geo-referencing process of the query image using the landmark recognition. If there are at least three points in the pattern image data with the associated latitude and longitude information, it is possible to perform the geo-referencing. In extreme case, if only one point in the pattern image had the latitude and longitude information, it would be possible to register the image if at least three landmarks would be recognized in the same frame.

Choosing the points that will have Latitude and Longitude information in the pattern image is an important aspect, since they are responsible for the system of linear equations for the image geo-referencing, and the system must be a possible and determined one. Those points will be the ground control points (GCP) for the process.

In terms of the Image geo-referencing, though, it is necessary to take into consideration the spatial resolution of the image when choosing the control points. The limitation on decimal representation in computers determines a minimum real distance that the points must have, in order not to affect the system of equations making it an impossible one. Moreover, the floating point precision must be taken into consideration when choosing each GCP.

The affine geometric transformation obtained is now used to find the corresponding points in the query image of each GCP of the pattern image. The points obtained are then used to obtain the image-to-image geo-referencing affine function, which will estimate the UAV position. We consider the center pixel of the query image as the point that maps the perspective center, since the images are taken in the nadir view, so that its estimated latitude and longitude is also considered the UAVs position.

3. EXPERIMENTS AND RESULTS

The experiments developed intended to validate the proposed method to estimate the position of an UAV during flight for a vision based navigation system. The tests performed focused on identifying position and on how accurate those positions were, compared with a previously known data.

There were two main tracks of experiments developed: Geo-referencing a Satellite Video frame and Auto-localization of a UAV. These different tests were built because of an initial lack of proper data to compare the results and validate the method. It is not easily found datasets of aerial images with corresponding flight data in the literature to compare results and in order to test the method and analyze the results this particular data was first produced using academic small quadcopter.

The experiments were performed in a MAC OSX 10.10 with a 2.6GHz Intel Core i7, 8GB 1600MHz DDR3 RAM and NVIDIA GeForce GT 650M 1024 MB, which could be used as a ground station for the UAV.

3.1 Geo-referencing a Satellite Video Frame

The first experiment was performed using a dataset provided by the IGRSS 2016 Data fusion Contest (DEIMOS, 2016). It consists of a Panchromatic data at 1m spatial resolution that was obtained from the DEIMOS-2 Satellite (CCD sensor) and a high definition video, with 1032 frames, and resolution of 3840x2160 pixels, acquired from the International Space Station at also 1m spatial resolution (CMOS sensor).

This test intended to observe three aspects. The first aspect was the use of satellite images, as they are a possible source of train images for landmarks. The second was related with the recognition in images provided by different sensors. Different sensors can capture different visual information and pose as a challenge for the recognition. The third and mas aspect was how the indirect image-to-image geo-referencing method would work.

The test then selected from the panchromatic data a landmark to be recognized in the video form the International Space Station.

Table 1. Indirect geo-referencing of the video frame, from landmark recognition.

Video Pos.	Image Pos.	Error (m)
(1711;600)	(1762;5057)	7.1
(2185;712)	(1957;4678)	6.6
(1675;1172)	(2219;5204)	16.5
(1172;1912)	(2714;5783)	25.8
(2841;1103)	(2426;4219)	3.9
(2751;464)	(1882;4151)	6.7
(1551;1033)	(2084;5273)	9.2
(335;1221)	(1953;6323)	33.7
(98;1732)	(2320;6633)	38.9
(2203;299)	(1622;4572)	4.0

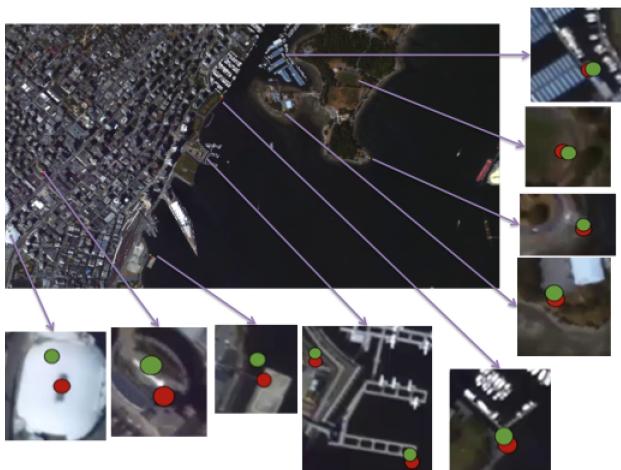


Figure 3. Result for the indirect Geo-referencing of a frame. The green dots are the real position and the red dots are the estimated position.

This selected landmark had latitude and longitude information in each pixels of the train image. From the recognition of the landmark, each frame of the video was then geo-referenced, using the method described in section 2.2.

Ten points from the video frame were randomly chosen in every part of the frame and used to evaluate the indirect geo-referencing method. It was decided to used ten points in order to evaluate how the error was distributed throughout the image. Their results on the indirect geo-referenced frame were compared with the corresponding points in the satellite geo-referenced image that provided the train image. Table 1 shows the results for these ten points.

The average error obtained is similar to the GPS error and the error that is usually obtained in image registration software, such as ENVI. Points with higher error are the ones that are more distant form the recognized landmark. A better registration would probably be obtained if more than one landmark were selected to be recognized in the frame. Figure 3 shows the result on the video frame.

3.2 UAV Auto-localization

The second experiment developed was closer to how would an aircraft recognize the landmark and estimate its position during flight. The dataset used for this experiment was a sequence of aerial images obtained from a Rotary-wing UAV (a quadcopter). The flight was performed in the 07/31/2015, at 16:30, average

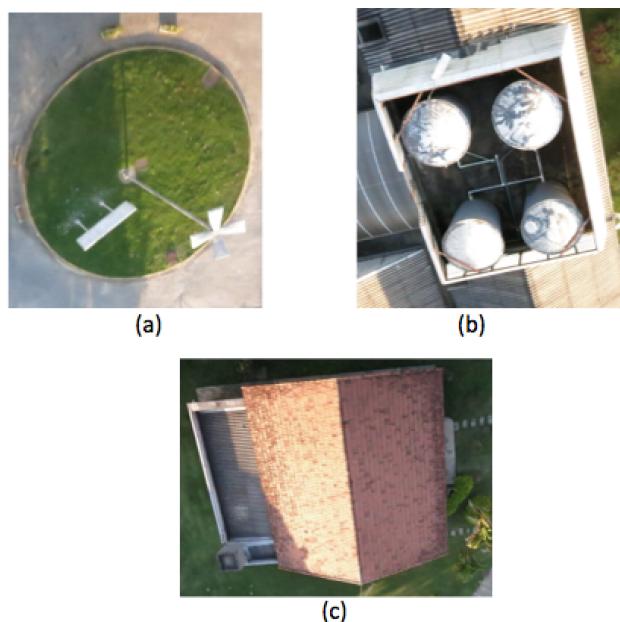


Figure 4. Landmarks chosen from the pattern image flight (07/31/2015). (a) is a roundabout landmark, (b) is the roof of the engineering building, and (c) is the roof of a house

flight height of 30m and average speed of 3m/s. The aerial images were taken with Nadir view, and at a frequency of 3 photos per second, with a pixel resolution of 4000x3000. It was also possible to obtain the flight logs, with information from the IMU and the GPS embedded in the aircraft, synchronized with each image from the sequences. The landmarks chosen were as in Figure 4, and they were obtained in the same aerial images. Four control points were randomly selected in each landmark, using a 2011 geo-referenced image of the area. These points are the GCP for the image registration, which is used for the position estimation.

The algorithm then was performed in the sequence of images. As the landmarks were recognized, the position was estimated and compared with the corresponding GPS data for each image in which the landmark was recognized. Table 2 shows the error in meters of the comparison between the estimated position and the position obtained by the onboard GPS. Figures 5, 6, and 7 illustrate the positions plotted on a map, and compared with the position of the center of each image were the landmark was recognized.

Table 2. Autolocalization error compared with the UAV GPS

Landmark	Error (m)
	19.8
	20.5
	24.4

The results have shown that estimations are inside the error radius (DRMS) of the data obtained by the GPS equipment embedded in the UAV. Moreover, the position from the Landmark recognition system seems to be more accurate than the GPS, when it is compared in a qualitative evaluation. When each aerial image in which the landmark was recognized is taken into consideration, if the GPS position and the estimated position are plotted in a map, it is possible to see that only in the estimated position the



Figure 5. Auto-localization position comparison between GPS, Landmark Recognition Estimation, Real UAV Position in the roundabout landmark.



Figure 6. Auto-localization position comparison between GPS, Landmark Recognition Estimation, Real UAV Position in the house landmark.

camera could capture a scene as the corresponding image is. In a more quantitative way, from the 2011 image of the area of flight, it was possible to geo-reference the UAV image using ENVI and estimate the central pixel latitude and longitude, which was considered as the UAV real position for comparison. From these results, it can be assumed that the method developed is a suitable alternative or redundant position estimator, to be used in a visual navigation system.

4. CONCLUSION

Landmark recognition-based autonomous navigation systems, then, proved to be a valid and promising strategy to be applied and further explored for a UAV visual navigation system. There are few works in this area and most of them are in a more heuristic approach, or use artificial landmarks. There are still experiments and adaptations to be developed in the recognition area, such as flight using other sensors and other environments. The auto-localization results although, were quite satisfactory, as they were more accurate than the GPS data available.

As Future works, an in-flight experiment is going to be per-



Figure 7. Auto-localization position comparison between GPS, Landmark Recognition Estimation, Real UAV Position in the rooftop landmark.

formed, in order to better evaluate the position estimated during flight. At first, the UAV will send its captured image to a ground station, which will process the image and obtain the position information (Latitude and Longitude) as soon as the landmark is recognized. The results will be compared with a usual GPS and a Referential GPS (RTK). Then, the algorithm will be adapted in order to be embedded in the aircraft.

5. ACKNOWLEDGEMENT

This research is part of a project funded by the Brazilian Air-Force. The Authors would like to thank FUNCATE and their Geoprocessing project for providing the means to participate in the conference.

REFERENCES

- Alcantarilla, P. F., Nuevo, J. and Bartoli, A., 2011. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell* 34(7), pp. 1281–1298.
- Azizi, A., Nourisola, H. and Ghiasi, A. R., 2016. 3d inertial algorithm of slam for using on uav. In: *Robotics and Mechatronics (ICROM), 2016 4th International Conference on*, IEEE, pp. 122–129.
- Blumenau, A., Ishak, A., Limone, B., Mintz, Z., Russell, C., Sudol, A., Linton, R., Lai, L., Padir, T. and Van Hook, R., 2013. Design and implementation of an intelligent portable aerial surveillance system (ipass). *Technologies for Practical Robot Applications (TePRA)* pp. 1–6.
- Braga, J. R., Velho, H. d. C. and Shiguemori, H., 2015. Estimation of uav position using lidar images for autonomous navigation over the ocean. In: *2015 9th International Conference on Sensing Technology (ICST)*, IEEE, pp. 811–816.
- Cruz, J. E. C., 2014. Reconhecimento de objetos em imagens orbitais com o uso de abordagens do tipo descriptor-classificador. Master's thesis, INPE.
- DeAngelo, M. P. and Horn, J. F., 2016. Landmark-aided localization for air vehicles using learned object detectors. *AIAA SciTech Proceedings*.

- DEIMOS, 2016. 2016 ieee grss data fusion contest. Online: <http://www.grss-ieee.org/community/technical-committees/data-fusion>.
- Dumble, S. and Gibbens, P., 2014. Efficient terrain-aided visual horizon based attitude estimation and localization. *Journal of Intelligent and Robotic Systems*.
- Farag, A. A. and Abdel-Hakim, A. E., 2004. Detection, categorization and recognition of road signs for autonomous navigation. In: *Proceedings of ACIVS*, pp. 125–130.
- Kezheng, L., Huanzhou, M. and Lang, W., 2015. Recognition algorithm of landmark for quadrotors aircraft based on image feature of corner points. In: *Information and Automation, 2015 IEEE International Conference on*, IEEE, pp. 1437–1440.
- Lee, J.-O., Kang, T., Lee, K.-H., Im, S. K. and Park, J., 2010. Vision-based indoor localization for unmanned aerial vehicles. *Journal of Aerospace Engineering* 24(3), pp. 373–377.
- LeMieux, J., 2012. Alternative uav navigation systems.
- Li, Y., Wang, S., Tian, Q. and Ding, X., 2015. A survey of recent advances in visual feature detection. *Neurocomputing* 149, pp. 736–751.
- Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60(2), pp. 91–110.
- Michaelsen, E. and Meadow, J., 2014. Stochastic reasoning for structural pattern recognition: An example from image-based uav navigation. *Pattern Recognition* 47(8), pp. 2732–2744.
- Muella, M. T. d. A. H., 2008. Morfologia e Dinâmica das Irregularidades de Pequena Escala e Imageamento Ionosférico por GPS. PhD thesis, INPE.
- Pajares, G., 2015. Overview and current status of remote sensing applications based on unmanned aerial vehicles (uavs). *Photogrammetric Engineering & Remote Sensing* 81(4), pp. 281–329.
- Quist, E. B. and Beard, R. W., 2016. Radar odometry on fixed-wing small unmanned aircraft. *IEEE Transactions on Aerospace and Electronic Systems* 52(1), pp. 396–410.
- Rady, S., Kandil, A. and Badreddin, E., 2011. A hybrid localization approach for UAV in GPS denied areas. *International Symposium on System Integration (SII)* pp. 1269–1274.
- Rublee, E., Garage, W., Park, M., Rabaud, V., Konolige, K. and Bradski, G., 2011. Orb: An efficient alternative to sift or surf. *International Conference on Computer Vision (ICCV)* pp. 2564 – 2571.
- Silva, C. A. O., 2015. Avaliação da técnica de casamento de imagens aplicada à localização geográfica de vants. Master's thesis, UFMG.
- Silva Filho, P. F. F., 2016. Automatic landmark recognition in aerial images for the autonomous navigation system of unmanned aerial vehicles. Master's thesis, Instituto Tecnológico de Aeronáutica (ITA), São José dos Campos-SP.
- Silva Filho, P., Rodrigues, M., Saotome, O. and Shiguemori, E. H., 2014. Fuzzy-based automatic landmark recognition in aerial images using orb for aerial auto-localization. In: *Advances in Visual Computing*, Springer, pp. 467–476.
- Singh, S. and Sujit, P., 2016. Landmarks based path planning for uavs in gps-denied areas this work is partially funded by fct grant sfrh/bpd/103962/2014. *IFAC-PapersOnLine* 49(1), pp. 396–400.
- Takahashi, H., Wrasse, C., Denardini, C., Costa, S., Ivo, A., Sant'Anna, N., Otsuka, Y. and Shiokawa, K., 2015. Large scale plasma bubbles observed by gps tec mapping over south america. In: *14th International Congress of the Brazilian Geophysical Society & EXPOGEF, Rio de Janeiro, Brazil, 3-6 August 2015*, Brazilian Geophysical Society, pp. 1437–1438.
- Zhao, X., Fei, Q. and Geng, Q., 2013. Vision based ground target tracking for rotor uav. *Control and Automation (ICCA)* pp. 1907 – 1911.
- Zhu, H. and Deng, L., 2016. A landmark-based navigation method for autonomous aircraft. *Optik-International Journal for Light and Electron Optics*.