VISION-BASED ATTITUDE DETERMINATION TO AID AIRBORNE NAVIGATION

João Matheus de Oliveira Arantes, jma@ita.br
Elder Moreira Hemerly, hemerly@ita.br
Technological Institute of Aeronautics, ITA-IEE-IEES
Electronics Division − Systems and Control Department
Pça. Mal. Eduardo Gomes, 50 − Vila das Acácias 12228-900 São José dos Campos − SP, Brazil

Abstract. *The use of unmanned aerial vehicles (UAV) requires robust, precise and autonomous navigation systems. In addition, these systems must be able to operate in a broad range of environments with a minimal error in order to successfully accomplish the missions for which they have been designed. A commonly used method is the well-known IMU/GNSS integration. Unfortunately, the GNSS data are not available all the time at any location, e.g., at urban or hostile environments. This fact motivates non-GNSS methods for aiding the inertial navigation system (INS), with the purpose of ensuring robust and accurate navigation solutions. Navigation techniques based on features of the environment have been demonstrated to be a promising approach in areas where GNSS can not provide satisfactory performance. Several algorithms can be used to extract features and match correspondence points in two image frames. A classical example is the Scale Invariant Features Transform (SIFT). However, in spite of SIFT robustness, it may return some false correspondences. Then, a technique such as Random Sample Consensus (RANSAC) must be applied in order to remove these outliers. Based on the consistent matches, relative orientation and position can be estimated by computing the relative motion of features between two or more consecutive image frames. This knowledge is relevant for reducing the drift errors of an onboard inertial navigation system (INS). The aim of this research is to develop a robust and efficient algorithm for tracking the locations of optical features in multiple images and estimate the relative rotation between them. The approach is based on Least-Square (LS) estimate of the projective transformation between two scenes. This estimate is refined further by employing Levenberg-Marquardt (LM) optimization algorithm.*

Keywords: *Airborne Navigation, Attitude Estimation, Homography Decomposition, Computer Vision*

1. INTRODUCTION

Unmanned Aerial Vehicles (UAVs) have been under a quick development in the last years. In general, unmanned systems are designed to offer two main advantages over manned aircraft: they are arguably inexpensive to procure, and also eliminate the risk to a pilot's life (Hwangbo, 2009). UAVs can be remote controlled aircrafts (e.g. flown by a pilot at a ground control station) or can fly autonomously based on pre-programmed flight plans or more complex dynamic automation systems. Traditionally, small UAVs have been used to perform tasks classified as dangerous or in missions where the presence of a pilot on board might be a limiting factor. Especially in the intelligence service, the unmanned aircrafts have been extensively used in surveillance, reconnaissance and tactical operations and safety (Perera *et al.*, 2006). In contrast to the recognized value of military applications, it is rare to encounter successful deployment of UAVs in civil applications. In order to insert such vehicles into commercial airspace, it is inherently important that these vehicles can generate collision-free motion plans and also be able to modify such plans during their execution in order to deal with contingencies which arise during the course of operation (Wzorek *et al.*, 2006).

These challenges can be overcome by developing critical capabilities for precise estimation of UAVs position and orientation. Usually, navigational methods such as GPS/INS coupling are used for these purposes. Unfortunately, standard techniques that rely heavily on GPS for navigation and control generally have difficulties in urban environments or hostile environments in which the satellite signals can be easily blocked, jammed or have their accuracy degraded (de Souza and Hemerly, 2009). In addition, accurate IMUs tend to be very expensive and the affordable ones typically exhibit low performance and can provide unbounded error results without GPS readings or other aiding sensors.

Based on these restrictions, researchers have investigated methods and sensors to accurately estimate vehicle pose when GPS is denied. A monocular camera system is useful due to its singular advantage to deliver multi-layered information in the format of images and has received growing interest as a local alternative/collaborative sensor to GPS based systems (Kaiser *et al.*, 2010). Contrary to GPS/INS that provides only information about the vehicle motion with respect to the inertial frame, vision can provide additional information relative to the environment, for example, how close the vehicle is to an obstacle, whether targets appear in the environment or how the vehicle is aligned with the horizon (Hwangbo, 2009).

Several works have demonstrated the effectiveness of recovering the attitude by using vision-based approaches. Kaiser *et al.* (2010) propose a geometric approach that creates a series of "daisy-chained' pose estimates. Veth and Raquet (2007) present an algorithm which integrates low-cost inertial and stereo imagery sensors to provide navigation solutions. Other works require some kind of a priori knowledge of the environment or landmarks such as Zhu *et al.* (2010), Sazdovski *et al.* (2010) and Daquan and Hongyue (2007).

In this paper, it is described an algorithm that can provide an estimate of the attitude of an unmanned aerial vehicle equipped with a single camera. The method is based on Least Square estimate of the homography between consecutive image frames. Differently from Kaiser *et al.* (2010), the problem of homography decomposition is solved by using the analytical method described in Malis and Vargas (2007). This closed form can improve the computational capabilities and also provide a better understanding on the homography decomposition for design the control schemes. Moreover, no *a prior* information concerning to the environment is required.

Section 2 describes details on problem scenario and methodology. Simulations are presented in Section 3 and Section 4 includes some experimental results. Finally, in Section 5 some conclusions are provided.

2. PROBLEM FORMULATION AND THE PROPOSED ALGORITHM

Consider an aerial vehicle equipped with a camera looking downwards. The objective of the technique proposed is to estimate the attitude of the UAV using camera data as a navigation aiding or as an alternative when the GPS signal is denied. As the camera has a limited field of view (FOV), the motion of the vehicle can cause observed feature points to leave the image. In this way, it is supposed that the system is capable of stop tracking a set of image features which is about to leave the FOV and begin tracking a new set of features in order to maintain the pose estimate. Bras *et al.* (2009) address this problem and propose a controller that explore camera's ability to undergo pan and tilt angular motions. Thus the estimation can continue indefinitely and is not limited by the camera's FOV (Kaiser *et al.*, 2010). Figure 1 illustrates the movement of the airship and how it acquires images of coplanar points in the world.

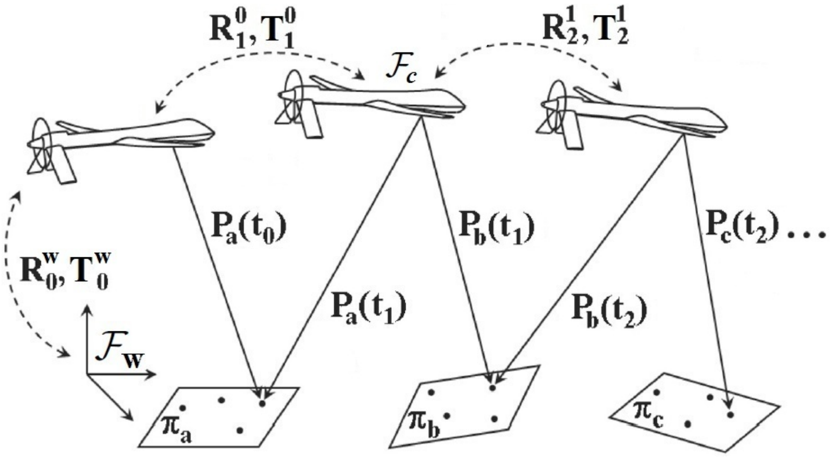


Figure 1. Illustration of image acquisition and pose estimation.

Without loss of generality, it is assumed that there are two coordinate frames: a body-fixed coordinate frame $\mathcal{F}_c$ that defines the *absolute* attitude $\mathbf{R_c^w}$ of the camera (and thus of the UAV) with respect to the constant world frame $\mathcal{F}_w$. It can also estimate the *relative* attitudes $\{\mathbf{R_1^0}, \mathbf{R_2^1}, ...\}$ of the camera between two consecutive images. Indeed, one gets the estimate of relative attitude based on the images and then calculate the absolute attitude from the previous data as

$$\mathbf{R_c^w} = \mathbf{R_0^w} \mathbf{R_1^0} \tag{1}$$

Still referring to Fig. 1, initially it is assumed that the UAV begins operating at time t_0, when the initial attitude $\mathbf{R_0^w}$ is known. Then an image of plane π_a is captured at this time, resulting the image points $\mathbf{p}_a(t_0)$ (in pixels). Next, at time t_1 another image of π_a is grabbed resulting the pixel coordinates $\mathbf{p}_a(t_1)$. As will be seen next, two images of the same scene are uniquely related by an homography. This transformation, by its turn, presents a rotation **R** and a translation **T** that can be obtained by homography decomposition. Thus, it is possible to recover the attitude of the camera from two corresponding images of the feature points $\mathbf{p}_a(t_0)$ and $\mathbf{p}_a(t_1)$. Moreover, the absolute attitude at a specific instant can be estimated by chaining the relative attitude estimates together without further use of the GPS.

Figure 2 gives an overview about this work and about the proposed method. When dealing with simulation analysis, it is necessary to generate input data to the algorithm. To that, it is used an open source toolbox designed to create and manipulate multiple camera scenarios (Mariottini and Prattichizzo, 2005). Furthermore, as a benchmark purpose, the results of the analytical homography decomposition method is compared to the SVD-based method proposed by Triggs (1998).

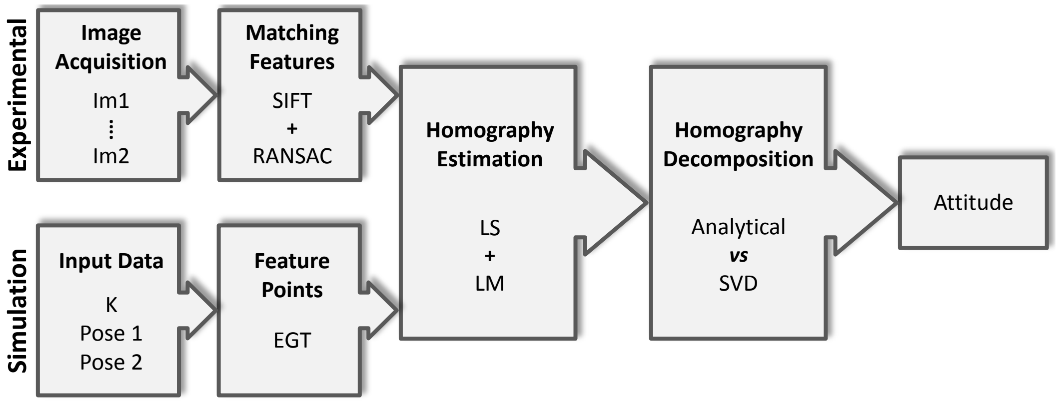


Figure 2. Overview of the proposed method regarding to simulation and experimental analysis.

2.1 Attitude estimate from two views

The attitude estimate is based on the homography decomposition method described in Malis and Vargas (2007). Consider a camera capturing images of a planar scene from two different locations a and b. As shown in Fig. 3(a), it can be assumed that there are two different camera frames $\mathcal{F}_a$ and $\mathcal{F}_b$, when camera pose is at time t and $t+1$, respectively. The 3D points in the plane π are projected on the image planes, so they have the normalized coordinates $\mathbf{m}_a = (x_a, y_a, 1)$ and $\mathbf{m}_b = (x_b, y_b, 1)$ with respect to the camera frames $\mathcal{F}_a$ and $\mathcal{F}_b$, respectively. The homogeneous image coordinates $\mathbf{p}_a = (u_a, v_a, 1)$ and $\mathbf{p}_b = (u_b, v_b, 1)$, in pixels, of each point can be obtained by applying the transformation in Eq. 2, where $\mathbf{K}$ is the upper triangular matrix containing the camera intrinsic parameters. In addition, there is a projective homography matrix $\mathbf{G}$ that transforms vector $\mathbf{p}_b$ into $\mathbf{p}_a$, up to a scale factor (Eq. 3). The camera frame is attached to the camera as shown in Fig. 3(b).

$$\mathbf{p} = \mathbf{K}\mathbf{m} \tag{2}$$

$$\alpha \mathbf{p}_a = \mathbf{G}\mathbf{p}_b \tag{3}$$

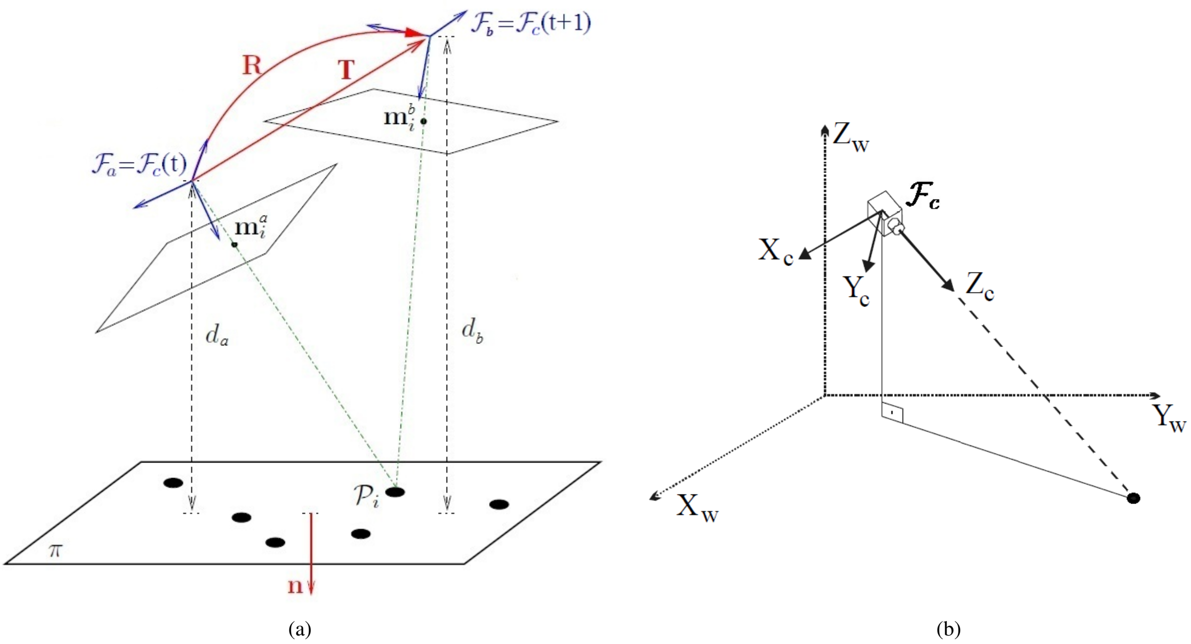


Figure 3. (a) Euclidean relationships between two camera poses; (b) camera coordinate frame.

The matrix $\mathbf{G}$ can be computed by matching several points from one image plane to another. But at least four points, provided that three of them are non-collinear, are required to determine a unique homography. This matrix is related to the transformation elements $\mathbf{R}$ and $\mathbf{T}$ and the normal to the plane $\mathbf{n}$ (this vector can be referred to both frames $\mathcal{F}_a$ or $\mathcal{F}_b$)

according to

$$\mathbf{G} = \gamma \mathbf{K}(\mathbf{R} + \mathbf{T}\mathbf{n}^T)\mathbf{K}^{-1} \tag{4}$$

The homography $\mathbf{H}$ in the Euclidean space can be computed from the projective homography $\mathbf{G}$ using the camera calibration matrix $\mathbf{K}$ as

$$\mathbf{H} = \frac{\mathbf{K}^{-1}\mathbf{G}\mathbf{K}}{\gamma} = \mathbf{R} + \mathbf{T}\mathbf{n}^T \tag{5}$$

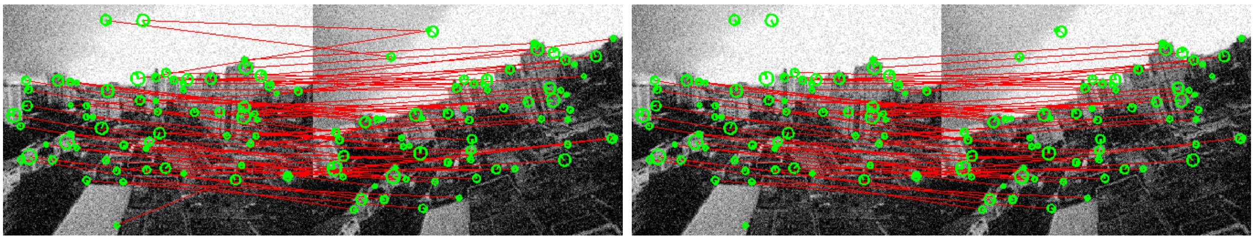
The problem of retrieving the elements $\mathbf{R}$, $\mathbf{T}$ and $\mathbf{n}$ from matrix $\mathbf{H}$, called Euclidean homography decomposition, is treated by several authors. Traditional methods based on SVD decomposition can be found in Faugeras and Lustman (1988), Triggs (1998) and Zhang and Hanson (1996). Although of the latter method is claimed to have closed-form expressions, the solutions are obtained numerically by using SVD decomposition. In contrast, Malis and Vargas (2007) present a really analytical solution for solving homography decomposition and provide the expressions of $\mathbf{R}$, $\mathbf{T}$ and $\mathbf{n}$ as functions of matrix $\mathbf{H}$. It allows to design control laws based on the relations among the possible solution.

All of the above decomposition methods generally return two physically possible solutions. Then, the correct one can be chosen by knowledge of the correct value for the normal vector or by using an image taken at a third pose (Kaiser *et al.*, 2010).

2.2 Image matching and homography estimate

As said before, in order to compute the matrix $\mathbf{G}$, several points have to be matched in both image planes. Thus, a robust and accurate method for extracting and matching distinctive invariant features from images must be applied. In computer vision, there is a well-known method to address this issue called Scale Invariant Feature Transform (SIFT) (Lowe, 2004). This technique is based on the distinctiveness of keypoints on each image. The keypoints are invariant to image rotation and scale and robust across a substantial range of affine distortion, addition of noise, and change in illumination. In this paper, it is is used an implementation of SIFT as a toolbox for MATLAB® developed by Vedaldi and Fulkerson (2008).

Although SIFT can provide robust results, some points can be mismatched and a refinement to remove these outliers is necessary. So Random Sample Consensus (RANSAC) is applied. RANSAC is an algorithm for robust fitting of models that was introduced by Fischler and Bolles (1981). The basic idea is, first of all, to randomly select a sample set of 4 non-collinear points (the minimal set required to compute the homography). Then, matrix $\mathbf{G}_i$ can be estimated using least-squares, as proposed in Murino *et al.* (2002). The distance $d_i = d(\mathbf{p}_{ai}, \mathbf{G}_i\mathbf{p}_{bi}) + d(\mathbf{p}_{bi}, \mathbf{G}_i^{-1}\mathbf{p}_{ai})$ is calculated and must be below a threshold to have a good model. The threshold is set to a rather low value ($t = 0.01$, corresponding to a couple of pixels), which possibly eliminates a few of the inliers, but at least it could be rather certain not to have any remaining outliers. Finally, a final LS fit is performed on the data points considered to be inliers.



(a) (b)

Figure 4. Application of SIFT and RANSAC for robust matching: (a) SIFT keypoints are matched with some outliers; (b) outliers removed after RANSAC processing.

Figure 4 illustrates how RANSAC can improve image matching. Two noisy image frames of the same scene from different views are correlated by the invariant features detected by SIFT. However, it can be seen in Fig. 4(a) that there are some outliers. In Fig. 4(b), after a RANSAC refinement, this outliers are removed.

After correctly mathcing the features in both images, RANSAC yields an homography matrix based on the LS embedded method. This homography is then used as an initial guess to refine the estimation using the LM optimization algorithm to minimize the symmetric transfer error (for all the inlier pairs), according to

$$d = \sum_i (d(\mathbf{p}_{ai}, \mathbf{G}\mathbf{p}_{bi})^2 + d(\mathbf{p}_{bi}, \mathbf{G}^{-1}\mathbf{p}_{ai})^2) \tag{6}$$

3. SIMULATION RESULTS

Without loss of generality, it was simulated a camera looking downwards moving along a trajectory, as if describing the motion of the aircraft fying at 50 m of height in a circular track with 20 m of diameter. To assist camera simulation, it was used the Epipolar Geometry Toolbox (EGT) provided by Mariottini and Prattichizzo (2005). Figure 5 shows the trajectory and some poses of the camera at different instants of time. In detail, it can be seen the camera at time *t* $\{X_{c0}, Y_{c0}, Z_{c0}\}$ ant at time *t+1* $\{X_{c1}, Y_{c1}, Z_{c1}\}$. It is also shown in Fig. 5 the set of 50 points in the world around the world-reference frame origin, whose pixel coordinates are computed by EGT. So, with a image acquisition rate of 15 FPS (frames per second), it is estimated the camera attitude between two instants of time.

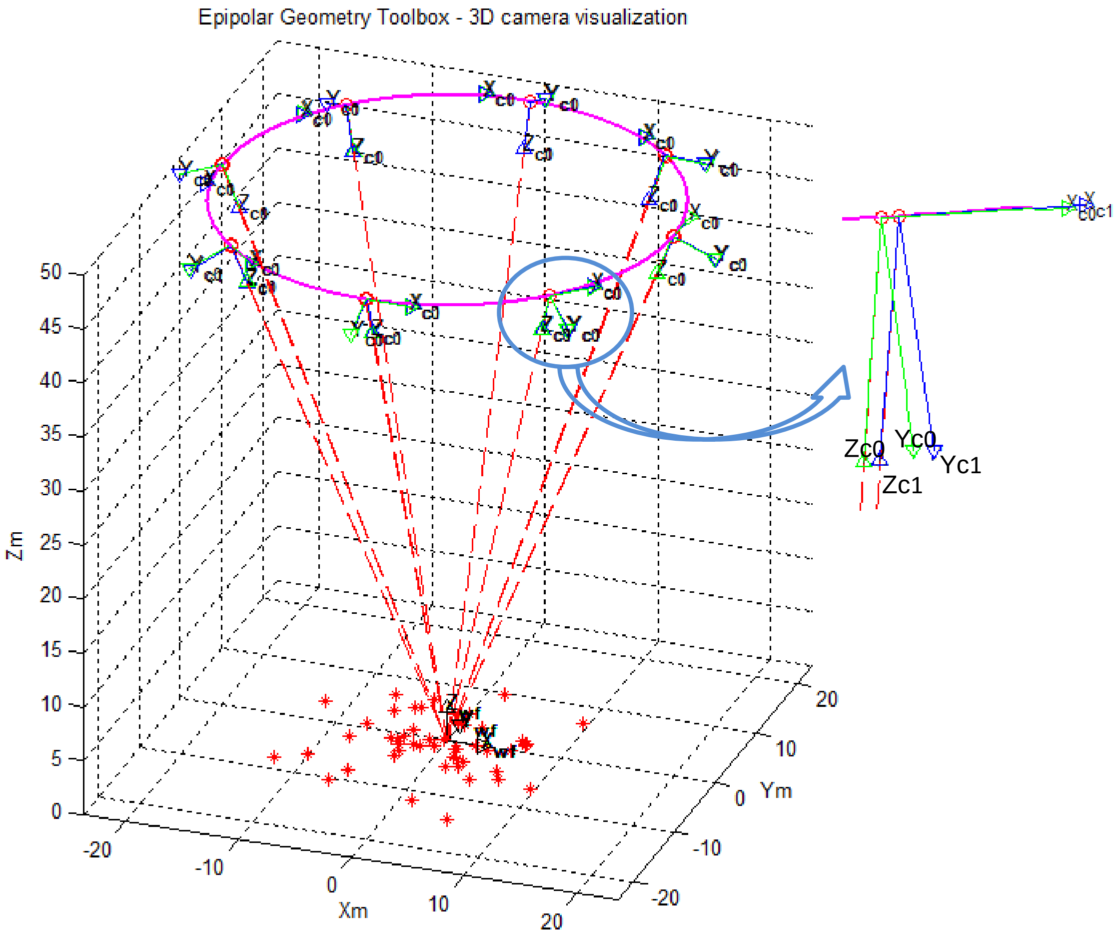


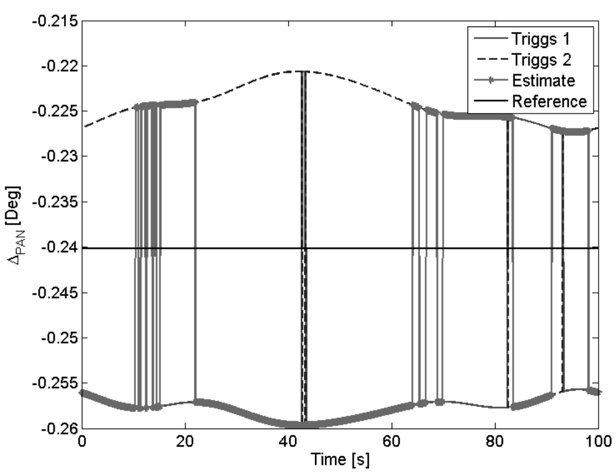
Figure 5. Camera trajectory and its positions at some instants.

The attitude is estimated based on the two solutions provided by Triggs method in order to compare them with those from the analytical method. These results are shown in Fig 6 and are similar. As expected the estimate toggles between two solutions since any general homography decomposition method provide two possible solutions. Moreover, as can be seen in Fig. 6(a), the estimate error is small and acceptable.

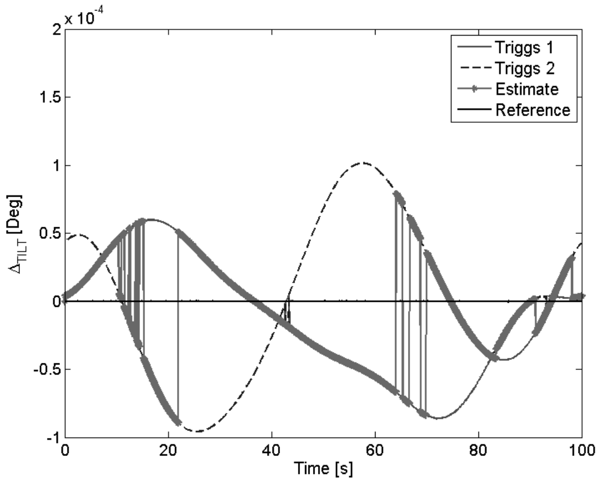
4. EXPERIMENTAL RESULTS

The experimental data have been taken using a Sony DSC-W320 digital camera. The intrinsic parameters of the camera was obtained using the camera calibration toolbox provided by Bouguet (2011). Initially, the camera was freely rotated in each axis without a precise measurement of the angles just to check if the algorithm was capable of a gross estimate of the rotations. The first rotation was applied in Z-axis (*Rz*), the second in X-axis (*Tilt*) and then in Y-axis (*Pan*). As can be seen in Fig. 7(a), the algorithm was capable of identify the rotations in each axis. It is important to stress that the estimated angles are between two consecutive poses/images. Therefore, this is the relative attitude of the camera. The absolute attitude for each frame in the trajectory was also computed and the results can be seen in Figure 7(b).

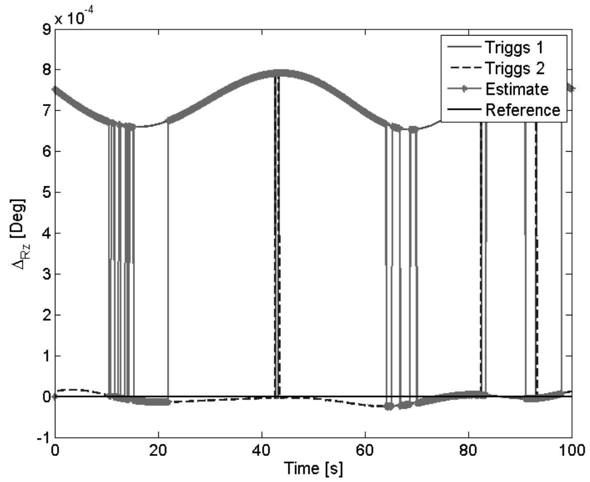
To evaluate the method accuracy, the camera was mounted on a platform so one could measure its rotation angle



(a)

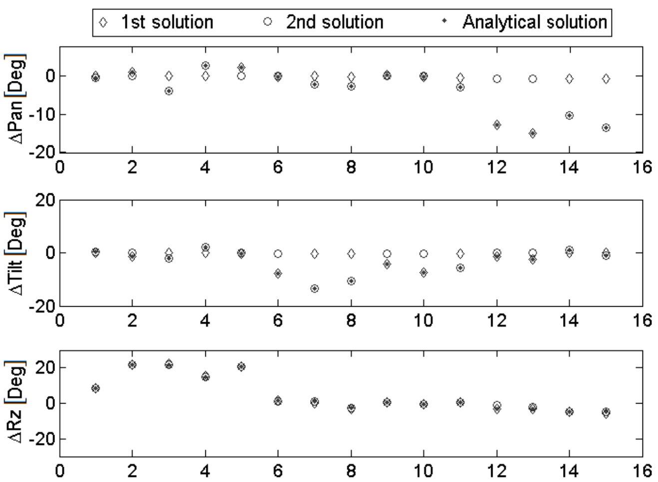


(b)

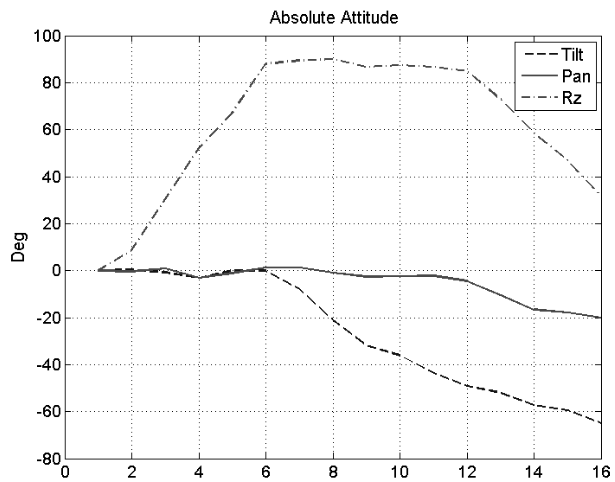


(c)

Figure 6. Attitude estimate over a circular trajectory.



(a)



(b)

Figure 7. Attitude estimate: (a) relative attitude between two positions (up: X-axis rotation; middle: Y-axis rotation; down: Z-axis rotation); (b) absolute attitude along the trajectory.

around a given axis. Each axis was rotated by approximately 20^o each time. Fig. 8 depicts the images before and after the rotations in X-axis (Fig. 8(a)), in Y-axis (Fig. 8(b)) and in Z-axis (Fig. 8(c)). The estimated attitudes are shown in Table 1, where there is a comparison of the results obtained by analytical decomposition method and the two solutions provided by Triggs' method. As expected, the possible solutions are identical and for the analytical method the normal vector estimate was imployed to select the valid solution. Gaussian noise was added to the images in order to evaluate the robustness of the method. Notice that with a noise variance of 25 some angles are estimated improperly. This occurs because the number of inliers in the images is near the low bound, just 5. Therefore, the homography estimates were poor causing the errors on the attitude estimates.

(a) (b) (c)

Figure 8. A rotation of 20^o around each axis: (a) X-axis rotation; (b) Y-axis rotation; (c) Z-axis rotation.

Table 1. Estimate of rotation around each axis, comparing the analytical results with Triggs results. Images were contaminated with additive Gaussian noise with zero mean and variance σ^2. Angles are measured in degrees.

	X-axis rotation			**Y-axis rotation**			**Z-axis rotation**		
σ^2	**S***	**Triggs 1**	**Triggs 2**	**S***	**Triggs 1**	**Triggs 2**	**S***	**Triggs 1**	**Triggs 2**
0	19.42	19.42	0.34	21.00	0.96	21.00	18.69	18.69	19.08
5	19.48	19.48	0.05	20.58	0.94	20.58	20.21	20.21	18.77
10	19.25	19.25	0.60	20.50	0.58	20.50	18.79	18.79	17.75
15	18.84	18.84	0.29	19.48	19.48	0.62	18.53	22.04	18.53
20	19.74	2.35	19.74	20.75	0.01	20.75	17.88	20.41	17.88
25	-26.37	-66.45	-26.37	18.94	18.94	-0.22	-54.86	-54.86	-32.56

S* stands for solution provided by analytical method.

5. CONCLUSIONS

The method described in this paper aims to determine the attitude of a camera that grabs images of coplanar points in the world. This information can eventually be used to estimate the attitude of an aircraft equipped with a single camera. An homography decomposition analytical method that provides the rotation between two images was applied.

By comparing the solutions of the analytical method with the two solutions of Triggs' method, see Tab. 1, and in other simulations, it was found that both methods provide similar solutions. However, the analytical method may have wider applications since it presents the advantage of being more easily implemented in real time, since no singular value decomposition is required.

The analytical method presents another benefit. As the two possible solutions are explicitly related, it is possible to design control strategies which employ the solutions, such as in Malis and Vargas (2007).

6. ACKNOWLEDGEMENTS

The authors would like to thank CNPq for the financial support.

7. REFERENCES

Bouguet, J.Y., 2011. "Camera calibration toolbox for matlab". URL `http://www.vision.caltech.edu/bouguetj/calib_doc/index.html`.

Bras, S., Cunha, R., Vasconcelos, J., Silvestre, C. and Oliveira, P., 2009. "Nonlinear attitude estimation using active vision and inertial measurements". In *48th IEEE Conference on Decision and Control (CDC)*. pp. 6496 –6501. ISSN 0191-2216. doi:10.1109/CDC.2009.5400904.

Daquan, T. and Hongyue, Z., 2007. "Vision based navigation algorithm for autonomic landing of uav without heading and attitude sensors". In *Signal-Image Technologies and Internet-Based System, 2007. SITIS '07. Third International IEEE Conference on Signal-Image Technologies and Internet-Based System*. pp. 972 –978. doi:10.1109/SITIS.2007.91.

de Souza, J.P. and Hemerly, E.M., 2009. "Vision-based control of fixed-wing unmanned aerial vehicle for autonomous road-following". In *Brazilian Symposium on Aerospace Eng. & Applications*.

Faugeras, O. and Lustman, F., 1988. "Motion and structure from motion in a piecewise planar environment". 0 RR-0856, INRIA. URL `http://hal.inria.fr/inria-00075698/en/`.

Fischler, M.A. and Bolles, R.C., 1981. "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography". *Communications of the ACM*, Vol. 24, No. 6, pp. 381–395.

Hwangbo, M., 2009. *Robust Monocular Vision-based Navigation for a Miniature Fixed-Wing Aircraft*. Ph.D. thesis, Robotics Institute − Carnegie Mellon University.

Kaiser, M.K., Gans, N.R. and Dixon, W.E., 2010. "Vision-based estimation for guidance, navigation, and control of an aerial vehicle". *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 46, No. 3, pp. 1064 –1077. ISSN 0018-9251. doi:10.1109/TAES.2010.5545174.

Lowe, D.G., 2004. "Distinctive image features from scale-invariant keypoints". *International Journal of Computer Vision*, Vol. 60, pp. 91–110. ISSN 0920-5691. URL `http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94`. 10.1023/B:VISI.0000029664.99615.94.

Malis, E. and Vargas, M., 2007. "Deeper understanding of the homography decomposition for vision-based control". Research Report RR-6303, INRIA. URL `http://hal.inria.fr/inria-00174036/en/`.

Mariottini, G. and Prattichizzo, D., 2005. "Egt: a toolbox for multiple view geometry and visual servoing". *IEEE Robotics and Automation Magazine*, Vol. 3, No. 12, pp. –. URL `http://egt.dii.unisi.it/`.

Murino, V., Castellani, U., Etrari, A. and Fusiello, A., 2002. "Registration of very time-distant aerial images". In *IEEE ICIP (3)*. pp. 989–992.

Perera, A.G.A., Srinivas, C., Hoogs, A., Brooksby, G. and Hu, W., 2006. "Multi-object tracking through simultaneous long occlusions and split-merge conditions". *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 666–673. ISSN 1063-6919. doi:http://doi.ieeecomputersociety.org/10.1109/CVPR.2006.195.

Sazdovski, V., Silson, P.M.G. and Tsourdos, A., 2010. "Attitude determination from single camera vector observations". In *5th IEEE International Conference Intelligent Systems (IS)*. pp. 49 –54. doi:10.1109/IS.2010.5548331.

Triggs, B., 1998. "Autocalibration from planar scenes". In *European Conference on Computer Vision*. Freiburg.

Vedaldi, A. and Fulkerson, B., 2008. "VLFeat: An open and portable library of computer vision algorithms". `http://www.vlfeat.org/`.

Veth, M.M. and Raquet, J., 2007. "Fusion of low-cost imaging and inertial sensors for navigation". Technical report, Air Force Institute of Technology.

Wzorek, M., Conte, G., Rudol, P., Merz, T., Duranti, S. and Doherty, P., 2006. "From motion planning to control - a navigation framework for an autonomous unmanned aerial vehicle". In *Proc. of the 21th Bristol International UAV Systems Conference*.

Zhang, Z. and Hanson, A.R., 1996. "3d reconstruction based on homography mapping". In *ARPA Image Understanding Workshop*. pp. 0249–6399.

Zhu, Z., Bhattacharya, S., Uijt de Haag, M. and Pelgrum, W., 2010. "Using single-camera geometry to perform gyro-free navigation and attitude determination". In *IEEE/ION Position, Location and Navigation Symposium (PLANS)*. pp. 858 –867. ISSN 2153-358X. doi:10.1109/PLANS.2010.5507188.

8. Responsibility notice

The author(s) is (are) the only responsible for the printed material included in this paper.