Position and Orientation of an Aerial Vehicle through Chained, Vision-Based Pose Reconstruction

K. Kaiser, N. Gans, W. Dixon

Dept. of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL, USA

While a Global Positioning System (GPS) is the most widely used sensor modality for aircraft navigation, researchers have been motivated to investigate other navigational sensor modalities because of the desire to operate in GPS denied environments. Due to advances in computer vision and control theory, monocular camera systems have received growing interest as an alternative/collaborative sensor to GPS systems. Cameras can act as navigational sensors by detecting and tracking feature points in an image. One limiting factor in this method is the current inability to relate feature points as they enter and leave the camera field of view. The contribution of this paper is a new vision-based state estimation method that allows sets of feature points to be related such that the aircraft position and orientation can be correlated to previous GPS data so that GPS-like navigation can be maintained in denied environments.

I. Introduction

GPS (Global Positioning System) is the primary navigational sensor modality used for vehicle guidance, navigation, and control. However, the frequently cited and highly comprehensive study referred to as the Volpe Report[1] indicates several vulnerabilities of GPS associated with signal disruptions. The Volpe Report delineates the sources of interference with the GPS signal into two categories, unintentional and deliberate disruptions. Some of the unintentional disruptions include ionosphere interference (also known as ionospheric scintillation) and radio frequency interference (broadcast television, VHF, cell phones, two-way pagers); whereas, some of the intentional disruptions involve jamming, spoofing, and meaconing. Some of the ultimate recommendations of this report were to, "create awareness among members of the domestic and global transportation community of the need for GPS backup systems..." and to "conduct a comprehensive analysis of GPS backup navigation..." which included ILS (Instrument Landing Systems), LORAN (LOng RAnge Navigation), and INS (Inertial Navigation Systems).[1]

The Volpe report acted as an impetus for many companies and institutions to investigate mitigation strategies for the vulnerabilities associated with the current GPS navigation aid protocol, nearly all following the suggested GPS backup methods that revert to the archaic/legacy methods. Unfortunately, these navigational modalities are limited by the range of their land-based transmitters, which are expensive and may not be feasible for remote or hazardous environments. Based on these restrictions, researchers have investigated local methods of estimating position when GPS is denied.

Given the advancements in computer vision and control theory, monocular camera systems have received growing interest as a local alternative/collaborative sensor to GPS systems. One issue that has inhibited the use of a vision system as a navigational aid is the difficulty in reconstructing inertial measurements from the projected image. Current approaches to estimating the aircraft state through a camera system utilize the motion of feature points in an image. Current approaches to recover the inertial state of the aircraft via a camera system include linear or nonlinear estimation methods. In contrast to these estimation methods, a geometric approach is proposed in this paper that uses a series of homography relationships. Specifically, a new method is proposed to create a series of daisy-chained images in which the feature points can be related so that the inertial coordinates of an aircraft can be determined between each successive image. Through these relationships, GPS data can be linked with the image data to provide inertial measurements in navigational regions where GPS is denied. Recently, a similar method using homography relationships between images to estimate the pose of an aircraft were presented by Caballero, et al.[2] Their method is limited to aircraft above a planar environment, while ours is applicable to piecewise planar landscapes.

Section II provides details on pose reconstruction using the epipolar homography. Section II also details the extension of pose estimation to include daisy-chaining multiple reference images. Simulations are presented in Section III to demonstrate the accuracy and effectiveness of this method. Section IV describes future steps necessary to improve this method for broader application.

II. Pose Reconstruction From Two Views

II.A. Euclidean Relationships

Consider a body-fixed coordinate frame $\mathcal{F}_c$ that defines the position and orientation of a camera with respect to a constant world frame $\mathcal{F}_w$. The world frame could represent a departure point, destination, or some other point of interest. The rotation and translation of $\mathcal{F}_c$ with respect to $\mathcal{F}_w$ is defined as $R(t) \in \mathbb{R}^{3\times 3}$ and $x(t) \in \mathbb{R}^3$, respectively. The camera rotation and translation from $\mathcal{F}_c(t_0)$ to $\mathcal{F}_c(t_1)$ between two sequential time instances, t_0 and t_1, is denoted by $R_{01}(t_1)$ and $x_{01}(t_1)$. During the camera motion, a collection of I (where $I \geq 4$) coplanar and non-colinear static feature points are assumed to be visible in a plane π. The assumption of four coplanar and non-colinear feature points is only required to simplify the subsequent analysis and is made without loss of generality. Image processing techniques can be used to select coplanar and non-colinear feature points within an image. However, if four coplanar target points are not available then the subsequent development can also exploit a variety of linear solutions for eight or more non-coplanar points (e.g., the classic eight points algorithm[3,4]), or nonlinear solutions for five or more points.[5]

A feature point $p_i(t)$ has coordinates $\bar{m}_i(t) = [x_i(t),\, y_i(t),\, z_i(t)]^T \in \mathbb{R}^3 \,\forall i \in \{1...I\}$ in $\mathcal{F}_c$. Standard geometric relationships can be applied to the coordinate systems depicted in Fig. 1 to develop the following relationships:

$$\begin{aligned} \bar{m}_i(t_1) &= R_{01}\bar{m}_i(t_0) + x \\ \bar{m}_i(t_1) &= H\bar{m}_i(t_0) \qquad (1) \\ \bar{m}_i(t_1) &= \left(R_{01}(t_1) + \frac{x_{01}(t_1)}{d(t_0)} n(t_0)^T\right)\bar{m}_i(t_0) \qquad (2) \end{aligned}$$

where $H(t)$ is the Euclidean homography matrix, and $n(t_0)$ is the constant unit vector normal to the plane π from $\mathcal{F}_c(t_0)$, and $d(t_0)$ is the constant distance between the plane π and $\mathcal{F}_c(t_0)$ along $n(t_0)$. After normalizing the Euclidean coordinates as

$$m_i(t) = \frac{\bar{m}_i(t)}{z_i(t)} \qquad (3)$$

(2) can be rewritten as

$$m_i(t_1) = \underbrace{\frac{z_i(t_0)}{z_i(t_1)}}_{\alpha_i} \underbrace{\left(R_{01}(t_1) + \frac{x_{01}(t_1)}{d(t_0)} n(t_0)^T\right)}_{H} m_i(t_0). \qquad (4)$$

where $\alpha_i \in \mathbb{R} \,\forall i \in \{1...I\}$ is a scaling factor.

II.B. Projective Relationships

Using the standard projective geometry, the Euclidean coordinate $\bar{m}_i(t)$ can be expressed in image-space pixel coordinates as $p_i(t) = [u_i(t),\, v_i(t),\, 1]^T$. The projected pixel coordinates are related to the normalized Euclidean coordinates, $m_i(t)$ by the pin-hole camera model as

$$p_i = Am_i, \qquad (5)$$

where A is an invertible, upper triangular camera calibration matrix[6] defined as

$$A \triangleq \begin{bmatrix} a & -a\cos\phi & u_0 \\ 0 & \frac{b}{\sin\phi} & v_0 \\ 0 & 0 & 1 \end{bmatrix}. \qquad (6)$$

In (6), u_0 and $v_0 \in \mathbb{R}$ denote the pixel coordinates of the principal point (the image center as defined by the intersection of the optical axis with the image plane), a, $b \in \mathbb{R}$ represent scaling factors of the pixel dimensions, and $\phi \in \mathbb{R}$ is the skew angle between camera axes.

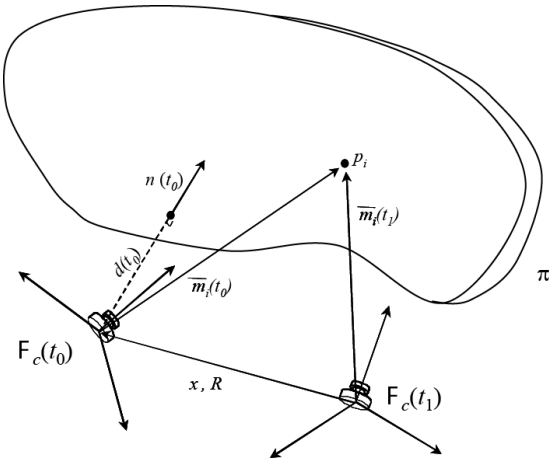

Figure 1. Euclidean relationships between two planar patches.

By using (5), the Euclidean relationship in (4) can be expressed as

$$\begin{aligned} p_i(t_1) &= \alpha_i AHA^{-1} p_i(t_0) \\ &= \alpha_i G p_i(t_0). \end{aligned} \tag{7}$$

Sets of linear equations can be developed from (7) to determine the projective and Euclidean Homography matrices $G(t)$ and $H(t)$ up to a scalar multiple. Given images of four or more feature points taken at $\mathcal{F}_c(t_0)$ and $\mathcal{F}_c(t_1)$, various techniques[7,8] can be used to decompose the Euclidean homography to obtain $\alpha_i(t_1)$, $n(t_0)$, $\frac{x_{01}(t_1)}{d(t_0)}$ and $R_{01}(t_1)$. The distance $d(t_0)$ must be separately measured (e.g., through an altimeter or radar range finder) or estimated using a priori knowledge of the relative feature point locations, stereoscopic cameras, or as an estimator signal in a feedback control.

II.C. Chained Pose Reconstruction for Aerial Vehicles

Consider an aerial vehicle equipped with a GPS and a camera capable of viewing a landscape. A technique is developed in this section to estimate the position and orientation using camera data when the GPS signal is denied. A camera has a limited field of view, and motion of a vehicle can cause observed feature points to leave the image. Therefore, this technique chains together pose estimations from sequential groups of points. This allows the estimation to continue when the camera's limited field of view would be inadequate.

The subsequent development assumes that the aerial vehicle begins operating in a GPS denied environment at time t_0, where the translation and rotation (i.e., $R_o(t_0)$ and $x_0(t_0)$ in Fig. 2) between $\mathcal{F}_c(t_0)$ and $\mathcal{F}_w(t_0)$ is known. The rotation between $\mathcal{F}_c(t_0)$ and $\mathcal{F}_w(t_0)$ can be determined through the bearing information of the GPS along with other sensors such as a gyroscope and/or compass. Without loss of generality, the GPS unit is assumed to be fixed to the origin of the aerial vehicle's coordinate frame, and the constant position and orientation of the camera frame is known with respect to the position and orientation of the aerial vehicle coordinate frame. This last assumption is necessary since the methods of Section II give the change in position and orientation of the camera and must be related to position and orientation of the vehicle through a coordinate transformation. The subsequent development further assumes that the GPS is capable of delivering altitude, perhaps in conjunction with an altimeter, so that the altitude $a(t_0)$ is known.

As illustrated in Fig. 2, the initial set of tracked coplanar and non-colinear feature points are contained in the plane π_a. These feature points have Euclidean coordinates $\bar{m}_{ai}(t_0) \in \mathbb{R}^3 \, \forall i \in \{1...I\}$ in $\mathcal{F}_c$. The plane π_a is perpendicular to the unit vector $n_a(t_0)$ in the camera frame, and lies at a distance $d_a(t_0)$ from the camera frame origin. At time t_1, the vehicle has some rotation $R_{01}(t_1)$ and translation $x_{01}(t_1)$ that can be determined from the images by decomposing the relationships given in (7). For notational simplicity, the subscript i is omitted in subsequent development.

As described in Section II, $R_{01}(t_1)$ and $\frac{x_{01}(t_1)}{d_a(t_0)}$ can be solved from two corresponding images of the feature points $p_a(t_0)$ and $p_a(t_1)$. A measurement or estimate for $d_a(t_0)$ is required to recover $x_{01}(t_1)$. This

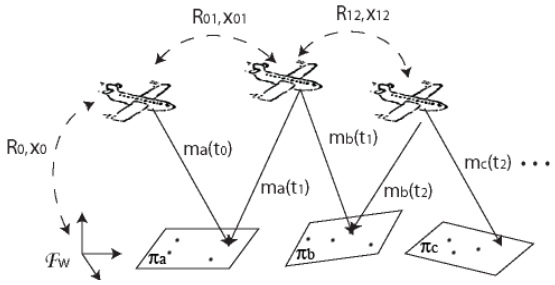

Figure 2. Illustration of pose estimation chaining

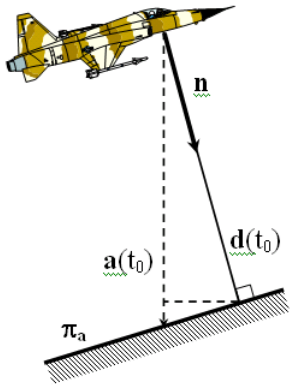

Figure 3. Illustration of depth estimation from altitude

estimation is possible with distance sensors or with a priori knowledge of the relative positions of the points in π_a. However, with an additional assumption, it is possible to estimate $d_a(t_0)$ geometrically using altitude information from the last GPS reading and/or an altimeter. From the illustration in Fig. 3, if $a(t_0)$ is the height above π_a (e.g. the slope of the ground is constant between the feature points and projection of the plane's location to the ground), then the distance $d_a(t_0)$ can be determined as

$$d_a(t_0) = n_a(t_0) \cdot a(t_0). \tag{8}$$

Once $R_{01}(t_1)$, $d_a(t_0)$, and $x_{01}(t_1)$ have been determined, the rotation $R_1(t_1)$ and translation $x_1(t_1)$ can be determined with respect to $\mathcal{F}_w$ as

$$\begin{aligned} R_1 &= R_0 R_{01} \\ x_1 &= R_{01} x_{01} + x_0. \end{aligned}$$

As illustrated in Fig. 2, a new collection of feature points $p_b(t)$ can be obtained that correspond to a collection of points on a planar patch denoted by π_b. At time t_2, the sets of points $p_b(t_1)$ and $p_b(t_2)$ can be used to determine $R_{12}(t_2)$ and $\frac{x_{12}(t_2)}{d_b(t_1)}$, which provides the rotation and scaled translation of $\mathcal{F}_c$ with respect to $\mathcal{F}_w$. If π_b and π_a are the same plane, then $d_b(t_1)$ can be determined as

$$d_b(t_1) = d_a(t_1) = d_a(t_0) + x_{01}(t_1) \cdot n(t_0). \tag{9}$$

When π_b and π_a are the same plane $x_{12}(t_2)$ can be correctly scaled, and $R_2(t_2)$ and $x_2(t_2)$ can be computed in a similar manner as described for $R_1(t_1)$ and $x_1(t_1)$. We can propagate estimations by chaining them together at each time instance without further use of GPS.

In the general case, p_b and p_a are not coplanar and (9) cannot be used to determine $d_b(t_1)$. If p_b and p_a are both visible for two or more frames, it is still possible to calculate $d_b(t)$ through geometric means. Take

t_{1-} as some time shortly before the daisy chain operation is performed, when both p_b and p_a are visible in the image. At time t_{1-}, we can solve an additional set of homography equations for the points p_b and p_a.at times t and t_{1-}

$$m_{ai}(t_1) = \underbrace{\frac{z_{ai}(t_{1-})}{z_{ai}(t)}}_{\alpha_a} \underbrace{\left(R + \frac{x}{d_a(t_{1-})} n_a(t_{1-})^T\right)}_{Ha} m_{ai}(t_{1-}) \tag{10}$$

$$m_{bi}(t_1) = \underbrace{\frac{z_{bi}(t_{1-})}{z_{bi}(t)}}_{\alpha_b} \underbrace{\left(R + \frac{x}{d_b(t_{1-})} n_b(t_{1-})^T\right)}_{Hb} m_{bi}(t_{1-}). \tag{11}$$

Note that in (10) and (11), R and x are the same, but the distance and normal to the plane are different for the two sets of points. The distance $d_a(t_{1-})$ is known from using (9). Define $x_b = \frac{x}{d_b(t_{1-})}$ and $x_a = \frac{x}{d_a(t_{1-})}$. The translation x is solved as

$$x = d_a(t_{1-}) x_a$$

and we can then find $d_b(t_{1-})$

$$d_b(t_{1-}) = \frac{x_b^T x}{\|x_b\|}.$$

$d_b(t_1)$ can then be found by using (9). Additional sensors, such as an altimeter, can provide an additional estimate in the change in altitude. This can be used in conjuction with (9) to update depth estimates.

III. Simulated Results

III.A. Position Estimation Through Camera Data

Several simulations are provided to test the performance of the proposed estimation scheme.

III.A.1. Position estimation using a single planar patch

The first simulation is focused on a single planar patch without the need to daisy-chain multiple planar patches. This scenario is useful to return the air vehicle to a possible GPS available location. The simulated camera is positioned above four coplanar points and moves in a circular path with constant linear velocity, altitude, and constant angular velocity in the camera frame, e.g. constant thrust and yaw, as depicted in Fig. 4.

At each time instant, the homography is calculated and the translation and rotation are determined. The position and orientation of the initial pose is known, including $d(t_0)$ and the initial distance to the plane containing the points. The position and rotation errors are presented (as roll-pitch-yaw angles) in Figure 5.

To investigate the effects of a poor estimate of $d(t_0)$, the simulation was repeated, but $d(t_0)$ was offset by 10%. The true and estimated trajectory are seen in Figure 6. The true trajectory is a solid line, while the estimate is shown as a dotted line. The estimation error is shown in Figure 7. The maximum error corresponds to a 10% error in the x direction and 4% in the y direction. As expected, rotation error is not affected by the error in estimating $d(t_0)$.

III.A.2. Position estimation by daisy-chaining multiple planar patches

The simulations in this section are limited to the ideal case that each planar patch is in the same plane. This assumption is valid for high altitude over relatively flat landscape. Future work will focus on the general case where the planar patches are not in the same plane. The camera now moves over three point patches and switches to the closest one at times $t = 50$ and $t = 80$. In Figure 8 the aircraft os shown to move in a straight path with constant velocity and a slight pitch angle. The pitch angle ensures that $d(t_i) \neq d(t_{i-1}), \forall i > 0$, and $d(t_1)$ must be estimated as in (9). Plots of the estimation errors in translation and rotation are seen in Figure 9.

A more complicated trajectory is shown in Figure 10. The trajectory given by the solid line is generated by a time varying linear velocity and a time-varying pitch and yaw angular velocity. Thus, at the switching

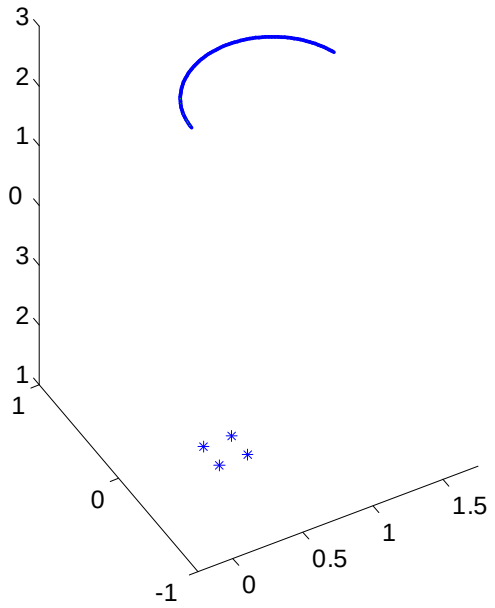

Figure 4. Circular trajectory above coplanar points

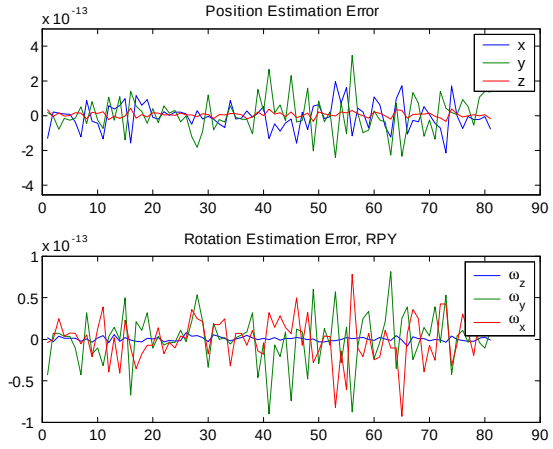

Figure 5. Estimation error for circular trajectory

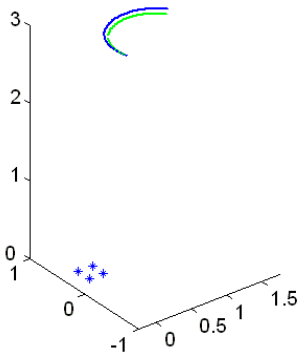

Figure 6. Circular trajectory with error in initial depth estimation

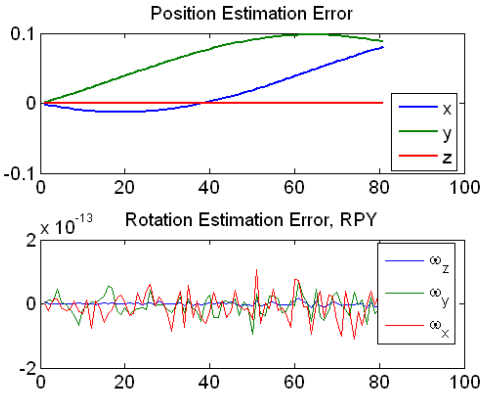

Figure 7. Estimation error for circular trajectory with error in initial depth estimation

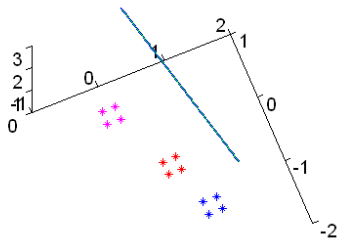

Figure 8. Daisy-chaining pose estimation for a straight trajectory

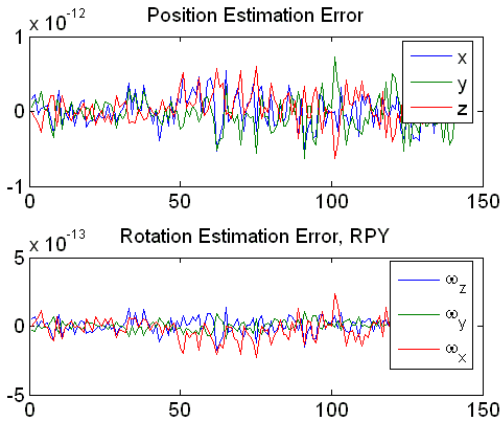

Figure 9. Estimation error for daisy-chaining along a straight trajectory

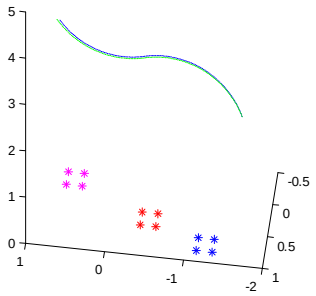

Figure 10. Daisy-chaining pose estimation for a six degree of freedom trajectory

times $t = 50$ and $t = 80$, $d(t_1)$ must be estimated as in (9). The estimated position is given by a dashed line, and some error develops over time for this trajectory. The translation and rotation errors are shown in Figure 11. Small errors in the position estimation arise from errors in estimating the translation from the homography $H(t)$, but the rotation error remains negligible.

IV. Future Work

The efforts in this paper describe the initial results for pose estimation of an aerial vehicle. Future efforts will target the estimate development when the assumptions are not satisfied. Of particular concern is the accuracy of the depth estimation, which scales the translation in (2). When the aerial vehicle cannot measure the altitude above the plane π_j, it is not possible to calculate the distance $d(t_0)$ as in (8). A scenario of this problem is when the aerial vehicle is over an environment with a changing topology. When the planes π_j are not the same, depth estimation cannot be propagated as in (9). There are several methods to estimate the depth. For example, comparing image velocity to true ground velocity of the camera can be used to estimate altitude. Knowledge of control inputs can be also be compared with the expected results in the image to estimate disturbances to the perfect model, such as wind speed. These estimations could be done through mathematical relationships, or preferably through the use of a state estimator that could be integrated directly with the flight controls of an aerial vehicle.

Our end goal is to enable flight in GPS denied environments. The Euclidean homography decomposition returns a coordinate transformation matrix which can be used to determine both Euler angles and body rates

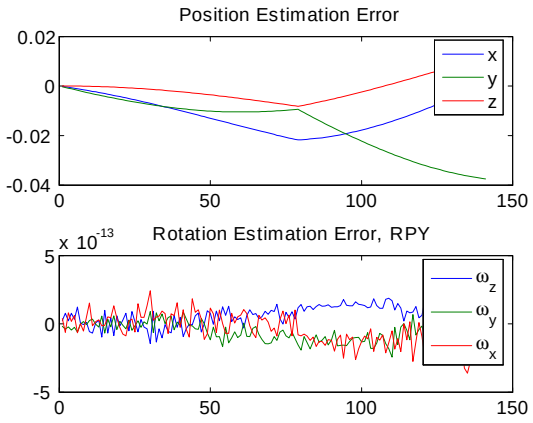

Figure 11. Estimation error for daisy-chaining along six degree of freedom trajectory

for use in a rudimentary autopilot design. Future work will also entail he use of the developed state estimate as a control strategy for an air vehicle, as well as addressing stability issues and uncertainty concerns such as noise, wind, and non-flat earth. Further investigation is warranted to see how the system performs under more complicated trajectories, particularly trajectories that obey flight dynamics.

V. Conclusions

GPS is a useful tool in estimating position, but it is not perfect. In the case that GPS is lost or jammed, a backup method of position estimation is needed. A rotation and translation estimation method is presented that uses the Euclidian homography between two camera views. Using the proposed methods, an aerial vehicle equipped with a camera could continue to estimate its position when GPS signals are not reliable or not available. A daisy chaining idea is proposed to include large motions where the camera's limited field of view would prove inadequate. Simulations of the position and orientation estimation are presented to illustrate the performance of the developed methods. Future work remains to extend the model to more realistic cases, and to integrate the position estimation and variation from the ideal model into the vehicle control framework.

VI. Acknowledgements

This research is supported in part by AFOSR contract number F49620-03-1-0381, AFRL contract number FA4819-05-D-0011, and by research grant No. US-3715-05 from BARD, the United States - Israel Binational Agricultural Research and Development Fund at the University of Florida.

References

[1]Center, J. A. V. N. T. S., "Vulnerability Assessment of the Transport Infrastructure Relying on the Global Positioning System," Report, Office of the Assistant Secretary for Transportation Policy, U.S. Department of Transportation, Aug. 2001.

[2]Caballero, F., Merino, L., Ferruz, J., and Ollero, A., "Improving vision-based planar motion estimation for unmanned aerial vehicles through online mosaicing," *2006, Proc. IEEE Conf. on Decision and Control*, 2006, pp. 2860– 2865.

[3]Longuet-Higgins, H., "A computer algorithm for reconstructing a scene from two projectionst," *Nature*, Sept. 1981, pp. 133 – 135.

[4]Hartley, R., *Computer Vision - ECCV'92, Lecture Notes in Computer Sciences*, Springer-Verlag, 1992.

[5]Nister, D., "An efficient solution to the five-point relative pose problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, June 2004, pp. 756– 770.

[6]Ma, Y., Soatto, S., Kosecká, J., and Sastry, S., *An Invitation to 3-D Vision*, Springer, 2004.

[7]Faugeras, O. and Lustman, F., "Motion and Structure from Motion in a Piecewise Planar Environment," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 2, No. 3, 1988, pp. 485–508.

[8]Zhang, Z. and Hanson, A., "reconstruction based on homography mapping," 1996, Zhang Z. and Hanson A.R. 3d reconstruction based on homography mapping. In ARPA Image Understanding Workshop, Palm Springs, CA, 1996.