2008 IEEE International Conference on
Robotics and Automation
Pasadena, CA, USA, May 19-23, 2008

Tracking Ground Targets with Measurements Obtained from a Single Monocular Camera Mounted on an Unmanned Aerial Vehicle

Dustin Deneault, Dale Schinstock, Chris Lewis

Abstract— **In this paper, a novel method is presented for tracking ground targets from an unmanned aerial vehicle (UAV) outfitted with a single monocular camera. The loss of observability resulting from the use of a single monocular camera is dealt with by constraining the target vehicle to follow ground terrain. An unscented Kalman filter (UKF) provides a simultaneous localization and mapping solution for the estimation of aircraft states and feature locations, which define the target's local environment. Another filter, a loosely coupled Kalman filter for the target states, receives 3D measurements of target position with estimated covariance obtained by an unscented transformation (UT). The UT uses the mean and covariance from the camera measurements and from the UKF estimated aircraft states and feature locations to determine the estimated target mean and covariance. Simulation results confirm the concepts developed.**

I. Introduction

MANY disciplines have investigated solutions to the target tracking problem for various applications including tracking military convoys, air traffic monitoring, and surveillance systems. Much of this work utilizes range and bearing sensors such as radar or laser scanners. Often theses sensors are located in stationary positions (i.e. fixed radar stations) [1]. However, small unmanned aerial vehicles (UAVs) are generally not equipped with either radar or laser range finders due to strict weight and power requirements. For power and weight limited systems, cameras are the sensor of choice. Such UAV mounted vision systems are suitable for geophysical surveying, remote sensing, ecological research, and autonomous navigation applications to name a few. This paper develops a method for tracking ground targets from a UAV using a single camera.

It is well known within the machine vision literature that scene reconstruction from images alone cannot resolve scale. Recovery of scale requires other sensors such as the global positioning system (GPS) that provide absolute position or a number of control points with known locations within the scene. Even with GPS sensors, this loss of scale continues to pose observability problems when tracking moving targets with video. Essentially, the component of the target's velocity along a line between it and the camera is ambiguous or unobservable. To overcome this limitation, at least for moving ground targets, this paper constrains the target to follow the terrain. Since terrain models are seldom available to the accuracy required, this paper borrows some techniques of simultaneous localization and mapping (SLAM) to estimate the terrain near the moving target. Unlike moving targets, stationary features are observable using vision from a moving platform.

The SLAM problem has received substantial attention in recent years. In general, researchers of SLAM techniques frame the problem as an optimal state estimation where the states include both the locations of map features and the navigation solution of the vehicle. A wide array of sensors and configurations of sensors have been studied including binocular vision, laser scanners, radar, and ultrasonic sensors. Recently, SLAM techniques have been applied to monocular vision systems in [2], [3], [12], and [13]. In [2], the authors assume the air vehicle maintains a constant altitude and that all the features are located in a flat plane. With these assumptions, they can reconstruct the 2D location of the aircraft and a 2D map of the features. In another work involving a UAV with monocular vision, [3], the researchers utilize artificially placed features with a known size. In this case, the range to features is estimated directly from the image, allowing a fully 3D reconstruction. A bearing only SLAM approach for UAVs is outlined in [13] where scale is resolved from an inertial measurement unit (IMU) sensor. This technique allows the development of a 3D map of the environment without knowledge of feature sizes. While this bearing only corrected SLAM implementation provides increased accuracy over an uncorrected inertial navigation solution, the aircraft position estimates contain significant drifting biases as compared to GPS aided inertial navigation solution (GPS/INS) estimates.

In this paper, both GPS and IMU data are used to overcome the scale issue and obtain the 3D estimate of both the UAV's navigation states and the terrain feature locations. Terrain features that were observed, but have gone out of view, are discarded in order to maintain reasonable computational loads. The moving target's states are then determined using a separate, loosely coupled state estimator that incorporates the SLAM estimates. An unscented transform (UT) is employed to propagate the means and uncertainties of the SLAM states and camera measurements into a 3D measurement vector with an estimated uncertainty.

D. Deneault is a Graduate Research Assistant for the Autonomous Vehicle Systems Lab at Kansas State University.

D. Schinstock Ph.D. is an Associate Professor for the Department of Mechanical and Nuclear Engineering at Kansas State University.

C. Lewis Ph.D. is an Associate Professor for the Department of Electrical and Computer Engineering at Kansas State University.

This paper is organized as follows. Section II begins by stating the fundamental assumptions necessary for this research. The next section outlines the state equation model development. Section IV follows with a brief heuristic explanation of the observability problems encountered in monocular vision approaches. Section V details the proposed solution, which utilizes the UT and loosely coupled estimators. Finally, simulated results are shown in Section VI, that illustrate the practicality of the theoretical development.

II. Assumptions

An integral step in the simulation process includes defining a set of realistic assumptions which accurately describe an experimental setup. In this paper, the setup will consist of modeling a relatively small UAV flying at low velocity and altitude over varying terrain. The UAV will obtain measurements from an IMU, a GPS estimator, and a monocular digital camera. The target properties are unknown. The following list details important assumptions used for model development.

- The aircraft's initial position defines the location of the inertial North East Down (NED) frame, $\{i\}$. This assumption is valid for highly maneuverable low velocity aircraft flying over a small area.
- The gravity vector will be constant during the duration of the flight. Its orientation will be aligned with the "down" axis of the inertial frame.
- The three angular rate sensors and three accelerometers within the IMU are located at the center of mass of the aircraft and are aligned with the axes of the body fixed reference frame, $\{b\}$. The GPS antenna is near the center of mass.
- The camera is located at the center of mass pointing down along the z axis of the body frame. Therefore, the camera frame is coincident with the body frame.
- The intrinsic parameters of the camera are known including the focal length, image center, skew coefficient, and radial and tangential distortions.
- A scale invariant distinctive feature based extraction algorithm exists for locating features within discrete images and for specifying the corresponding image coordinate locations from overlapping images.
- The targets are limited to ground targets that must remain in contact with local ground terrain.

III. Model Development

As is typical for GPS aided inertial navigation solutions (GPS/INS), kinematic equations rather than dynamic equations are used. This allows for estimation of the aircraft and target states without knowledge of the intrinsic parameters of the aircraft and target. The resulting state estimation technique is less application specific. Also, kinematic relationships generally tend to decrease the number of states necessary for estimation while providing a quantifiable means for determining input and measurement noise.

For our development, three sets of kinematic equations are required: one for the UAV, one for the features, and one for the target. The general form for these nonlinear state equations is given by

$$\dot{\vec{x}} = \vec{f}(\vec{x}, \vec{u}, \vec{w}) \tag{1}$$

$$\vec{z} = \vec{h}(\vec{x}, \vec{v}) \,. \tag{2}$$

where $\vec{x}$ is the state vector, $\vec{u}$ is the input vector, $\vec{w}$ is the zero mean white process noise vector, $\vec{v}$ is the zero mean white noise output vector, and $\vec{z}$ is the output vector. In the following discussion, (1) will be referred to as the dynamic model as it defines the relationship between the time derivative of the state vector and the nonlinear vector function, $\vec{f}(\cdot)$, of the states, inputs, and input noise. Equation (2) will be referred to as the observation or output model where the nonlinear output vector function, $\vec{h}(\cdot)$, defines the combination of states and output noise that form the output vector.

A. Aircraft Nonlinear State Equations

The definitions for the aircraft state vector, input vector, and output vector are shown below. The "a" subscript denotes aircraft vectors while the subscript within braces denotes the inertial or body fixed reference frame. Measurement vectors are augmented with an additional "m" subscript.

$$\vec{x}_a \equiv \begin{bmatrix} \vec{p}_{a\{i\}}^T & \dot{\vec{p}}_{a\{i\}}^T & \vec{q}^T \end{bmatrix}^T \tag{3}$$

$$\vec{u}_a \equiv \begin{bmatrix} \vec{a}_{m\{b\}}^T & \vec{\omega}_{m\{b\}}^T \end{bmatrix}^T \tag{4}$$

$$\vec{z}_a \equiv \begin{bmatrix} \vec{p}_{am\{i\}}^T & \dot{\vec{p}}_{am\{i\}}^T \end{bmatrix}^T \tag{5}$$

The state vector, $\vec{x}_a$, consists of the UAV's inertial position vector, $\vec{p}_{a\{i\}}$, inertial velocity vector, $\dot{\vec{p}}_{a\{i\}}$, and the quaternion vector, $\vec{q}$, where

$$\vec{q} = \begin{bmatrix} q_0 & q_1 & q_2 & q_3 \end{bmatrix}^T \,. \tag{6}$$

The input vector, $\vec{u}_a$, contains the body axes measurements from the IMU accelerometers, $\vec{a}_{m\{b\}}$, and gyros, $\vec{\omega}_{m\{b\}}$. The output, $\vec{z}_a$, contains GPS measurements for the inertial position vector, $\vec{p}_{am\{i\}}$, and inertial velocity vector, $\dot{\vec{p}}_{am\{i\}}$.

1) Aircraft Dynamic Equations

The time derivatives of the states, $\dot{\vec{x}}_a$, are given by the following equations.

$$\frac{d}{dt}(\vec{p}_{a\{i\}}) = \dot{\vec{p}}_{a\{i\}} \tag{7}$$

$$\ddot{\vec{p}}_{a\{i\}} = \mathbf{R}_{\{ib\}} \vec{a}_{m\{b\}} + \mathbf{R}_{\{ib\}} \vec{w}_{a\{b\}} + \begin{bmatrix} 0 & 0 & g \end{bmatrix}^T \tag{8}$$

$$\dot{\bar{q}} = \frac{1}{2}\mathbf{\Omega}(q)(\bar{\omega}_{m\{b\}} + \bar{w}_{\omega\{b\}}) \tag{9}$$

where

$$\mathbf{\Omega}(q) = \begin{bmatrix} -q_1 & -q_2 & -q_3 \\ q_0 & -q_3 & q_2 \\ q_3 & q_0 & -q_1 \\ -q_2 & q_1 & q_0 \end{bmatrix} \tag{10}$$

The time derivative of position within the inertial frame, $\dot{\bar{p}}_{a\{i\}}$, is simply the inertial velocity vector as expressed in (7). Equation (8) defines the time rate change of velocity in the inertial frame, $\ddot{\bar{p}}_{a(i)}$, which must be equal to the kinematic accelerations within the inertial frame. The accelerometers located on the body frame measure the true accelerations corrupted with components of zero mean white noise, $\bar{w}_{a\{b\}}$, and the constant magnitude gravity vector, g. Therefore, the true accelerations are obtained by eliminating the gravity vector component and rotating the acceleration measurement into the correct frame with the body frame to inertial frame rotation matrix, $\mathbf{R}_{\{ib\}}$. The next two equations, (9) and (10), are commonly referred to as the "strapdown equations" as they represent the relationship between the time derivatives of the quaternion states, whose elements are functions of the rotation matrix, and the gyro measurements [6]. As with the accelerometer measurements, the gyro measurements, $\bar{\omega}_{m\{b\}}$, are also corrupted with noise, $\bar{w}_{\omega\{b\}}$, as shown in (9).

2) Aircraft Observation Equations

The output given by the GPS includes estimates of both inertial position, $\bar{p}_{am\{i\}}$, and inertial velocity measurements, $\dot{\bar{p}}_{am\{i\}}$. Both of these measurement vectors contain the true inertial position and velocity and their corresponding noise components, $\bar{v}_{p_a\{i\}}$ and $\bar{v}_{\dot{p}_a\{i\}}$.

$$\bar{z}_a = \begin{bmatrix} \bar{p}_{am\{i\}} \\ \dot{\bar{p}}_{am\{i\}} \end{bmatrix} = \begin{bmatrix} \bar{p}_{a\{i\}} \\ \dot{\bar{p}}_{a\{i\}} \end{bmatrix} + \begin{bmatrix} \bar{v}_{p_a\{i\}} \\ \bar{v}_{\dot{p}_a\{i\}} \end{bmatrix} \tag{11}$$

B. Feature Nonlinear State Equations

The following definitions list the state vector, input vector, and output vector associated with the features. The "f" subscript denotes feature vectors.

$$\bar{x}_f \equiv \begin{bmatrix} \left(\bar{p}^1_{f\{i\}}\right)^T & \cdots & \left(\bar{p}^{N_f}_{f\{i\}}\right)^T \end{bmatrix}^T \tag{12}$$

$$\bar{u}_f \equiv \begin{bmatrix} \bar{0}^T & \cdots & \bar{0}^T \end{bmatrix}^T \tag{13}$$

$$\bar{z}_f \equiv \begin{bmatrix} \left(\bar{i}^1_{fm\{b\}}\right)^T & \cdots & \left(\bar{i}^{N_f}_{fm\{b\}}\right)^T \end{bmatrix}^T \tag{14}$$

The feature state vector, $\bar{x}_f$, in (12) contains the feature vector locations, $\bar{p}_{f\{i\}}$, relative to the NED inertial frame. The number of estimated features may vary with time as new features come into observation and other features go out of observation. The superscripts are used to distinguish the various features where N_f denotes the total number of features comprising the state vector at any given time. The variable size input vector, $\bar{u}_f$, in (13) is all zero vectors since no input is required for the feature dynamic equations. The output vector, $\bar{z}_f$, contains the image space vector measurements, $\bar{i}_{fm\{b\}}$, for each feature. Each $\bar{i}_{fm\{b\}}$ consists of two elements which define the measured x and y components of a single image space feature representation relative to $\{b\}$.

1) Feature Dynamic Equations

Features are assumed stationary, which implies the time derivative of the feature state vector, $\dot{\bar{x}}_f$, is zero.

$$\dot{\bar{p}}_{f\{i\}} = \bar{0} \tag{15}$$

2) Feature Observation Equations

A monocular camera onboard the aircraft captures images at periodic intervals. The observation equations determine the image coordinates of the features using aircraft position and orientation, the intrinsic camera parameters, and the feature locations. Fig. 1 depicts relationships between the inertial frame, the camera frame, and a feature viewed through a pinhole camera model.

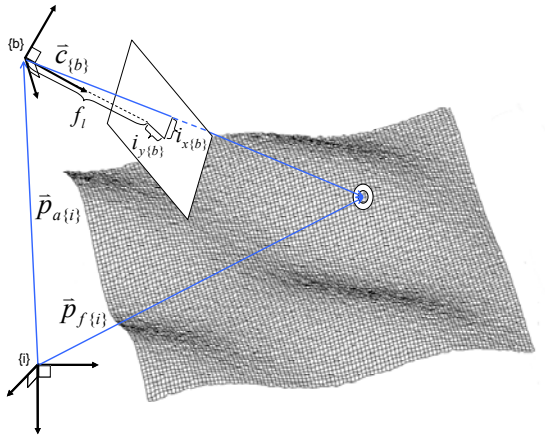

Figure 1: Feature observation development assuming pinhole camera model..

The vector, $\bar{c}_{\{b\}}$, from the camera frame to the feature, is given by the following equation

$$\bar{c}_{\{b\}} = \mathbf{R}_{\{bi\}}(\bar{p}_{f\{i\}} - \bar{p}_{a\{i\}}). \tag{16}$$

where the rotation matrix, $\mathbf{R}_{\{bi\}}$, transforms the inertial frame representations of the feature and aircraft into the body frame. The image space feature coordinates are realized when the z component of $\bar{c}_{\{b\}}$ is equal to the focal length, f_l, of the camera. Therefore, the actual image coordinates $i_{x\{b\}}$ and $i_{y\{b\}}$ of a feature are computed as

follows.

$$\vec{i}_{f\{b\}} = \begin{bmatrix} i_{x\{b\}} \\ i_{y\{b\}} \end{bmatrix} = \frac{f_l}{\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \vec{c}_{\{b\}}} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \vec{c}_{\{b\}} \quad (17)$$

The measured image space coordinate vectors, $\vec{i}_{fm\{b\}}$, consist of the true image coordinate vector, $\vec{i}_{f\{b\}}$, plus a noise vector, $\vec{v}_{i_f\{b\}}$, due to camera calibration and feature extraction errors.

$$\vec{z}_f = \begin{bmatrix} \vec{i}^{\,1}_{fm\{b\}} \\ \vdots \\ \vec{i}^{\,N_f}_{fm\{b\}} \end{bmatrix} = \begin{bmatrix} \vec{i}^{\,1}_{f\{b\}} \\ \vdots \\ \vec{i}^{\,N_f}_{f\{b\}} \end{bmatrix} + \begin{bmatrix} \vec{v}^{\,1}_{i_f\{b\}} \\ \vdots \\ \vec{v}^{\,N_f}_{i_f\{b\}} \end{bmatrix} \quad (18)$$

C. Target Nonlinear State Equations

Like the aircraft states, the target states will consist of kinematic relationships. However, unlike the aircraft model, no direct acceleration measurements are available from the target. One commonly used motion model assumes the velocity of the target is a random walk with acceleration modeled using zero mean Gaussian white noise. For this technique, a larger standard deviation in acceleration noise corresponds to greater uncertainty in the target states thereby weighting the observation more heavily than the prediction in the state estimation process.

Modeling the target velocity as a random walk with unchanging noise statistics and dynamic model may not be adequate for highly maneuverable targets. The Interacting Multiple Model (IMM) estimator can provide improved accuracy for modeling such targets at the expense of increased computational complexity. The interested reader should consult [1], [7], and [8]. The target tracking concepts developed in this paper pertain equally to both the IMM target estimator and the basic single target model estimator used for the remainder of this development.

The following equations represent the target state vector, input vector, and output vector. The "t" subscript denotes target vectors.

$$\vec{x}_t \equiv \begin{bmatrix} \vec{p}^{\,T}_{t\{i\}} & \dot{\vec{p}}^{\,T}_{t\{i\}} \end{bmatrix}^T \quad (19)$$

$$\vec{u}_t \equiv \vec{w}_{t\{i\}} \quad (20)$$

$$\vec{z}_t \equiv \vec{i}_{tm\{b\}} \quad (21)$$

In (19), the target state vector, $\vec{x}_t$, contains the target position vector, $\vec{p}_{t\{i\}}$, and velocity vector, $\dot{\vec{p}}_{t\{i\}}$, relative to the inertial frame. Since the velocity is modeled using a random walk, the input vector for the target model, $\vec{u}_t$, is a noise term, $\vec{w}_{t\{i\}}$. The target output vector, $\vec{z}_t$, consists of the camera sensor measurement vector, $\vec{i}_{tm\{b\}}$, which contains the two image space coordinate measurements for the target.

1) Target Dynamic Equations

The time derivative of the target state vector, $\dot{\vec{x}}_t$, is a concatenation of the velocity and acceleration terms.

$$\frac{d}{dt}(\vec{p}_{t\{i\}}) = \dot{\vec{p}}_{t\{i\}} \quad (22)$$

$$\ddot{\vec{p}}_{t\{i\}} = \vec{w}_{t\{i\}} \quad (23)$$

2) Target Observation Equations

The target observation model is very similar to the feature observation model since the camera image provides the only output measurement for the target. The image space coordinate vector, $\vec{i}_{t\{b\}}$, is given by (17) when $\vec{p}_{f\{i\}}$ in (16) is replaced with $\vec{p}_{t\{i\}}$. The target observation model consists of the true image space vector, $\vec{i}_{t\{b\}}$, and a noise vector, $\vec{v}_{i_t\{b\}}$, for similar reasons as described for the feature observation equations.

$$\vec{z}_t = \vec{i}_{tm\{b\}} = \vec{i}_{t\{b\}} + \vec{v}_{i_t\{b\}} \quad (24)$$

IV. Observability Problems

This section discusses the observability of the state equations developed for the aircraft, feature, and target state representations. A set of state equations is said to be observable when knowledge of the input and output over a finite time interval allows unique determination of the initial state vector [9]. A traditional test for linear systems analyzes the rank of the observability matrix. If the rank of the observability matrix is greater than or equal to the number of states, the system is considered observable. Observability is the fundamental requirement for any observer, as the loss of observability corresponds to the inability to generate converging state estimates.

Nonlinear equations, unfortunately, have to be linearized before the observability matrix can be formed. The analytical observability matrix developed from the first order term of the Taylor series expansion of the nonlinear equations contains elements dependent upon current states and inputs. Therefore, in certain instances, observability may be lost for some states depending upon the current operating point. The process of determining rank from linearized nonlinear systems results in a complicated matrix and sometimes prohibits physical intuition of observability loss conditions. This paper provides an intuitive observability explanation based upon heuristic arguments developed during simulations.

A. Feature Observability

The target tracking solution requires information obtained from the 3D location of features, and therefore, their observability is analyzed here. As the camera poses are consistent with the position and orientation of the aircraft, the state equations must include both the aircraft and feature states, thus defining the SLAM problem.

$$\dot{\vec{x}} = \begin{bmatrix} \dot{\vec{x}}_a^T & \dot{\vec{x}}_f^T \end{bmatrix}^T \quad (25)$$

$$\vec{z} = \begin{bmatrix} \vec{z}_a^T & \vec{z}_f^T \end{bmatrix}^T \quad (26)$$

Fig. 2 illustrates the relationships between the aircraft states and feature states through multiple observations. At time $(k-1)\Delta t$, the UAV, denoted with a solid outline, flies directly over a stationary feature and captures an image.

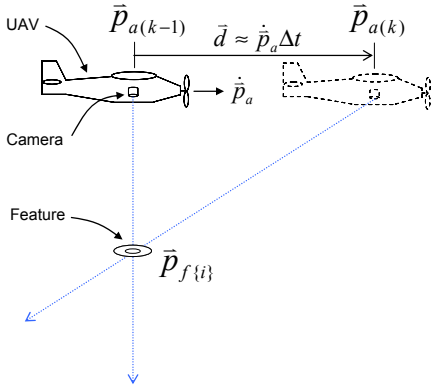

Figure 2: The triangulation of a feature in space as viewed from two different poses.

The vector pointing from the camera through the actual feature location denotes the line in space corresponding to the family of possible, but still unknown, 3D feature locations. A short time later, Δt, the UAV, now denoted by its dashed outline, has traveled a small distance and captures another image. A new vector, pointing from the new camera location to the feature, symbolizes the line along which the feature should reside. Assuming the feature's location did not change, measurement of these two vectors fully triangulates the feature's location. However, the scale of the resulting triangle remains unknown. The scale of this triangle is determined from the distance vector, $\vec{d}$, pointing from the previous aircraft position to the current position, which is predicted by the integration of the IMU measurements and observed through GPS. Realistically, the measurement of these two vectors is corrupted by noise and uncertainty among the state estimates, resulting in skew vectors with some residual error in the feature's location. With sufficient measurements taken from different aircraft positions and orientations, the mean location of the feature in 3D space can be estimated accurately. The camera, and hence, the aircraft must move to allow triangulations and residual corrections to exist.

B. Target Observability

The state equations necessary for the examination of observability for the target tracking model must contain both the aircraft and target state equations as the target measurement depends upon the aircraft states in much the same way as the feature observations did.

$$\dot{\vec{x}} = \begin{bmatrix} \dot{\vec{x}}_a^T & \dot{\vec{x}}_t^T \end{bmatrix}^T \quad (27)$$

$$\vec{z} = \begin{bmatrix} \vec{z}_a^T & \vec{z}_t^T \end{bmatrix}^T \quad (28)$$

The major difference between target estimation and feature estimation originates from the dynamic nature of the target. This is illustrated in Fig. 3.

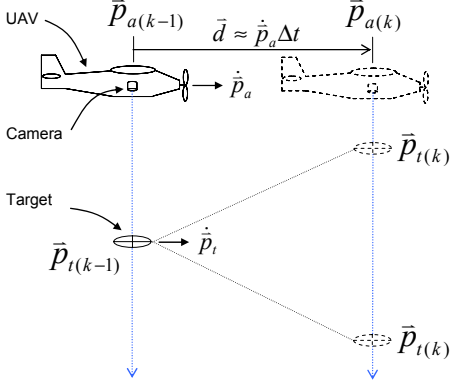

Figure 3: The lack of triangulation of a moving target in space as viewed from two different poses.

Once again, the vectors from the camera passing through the target represent the families of possible target locations. The target, in this case, has a velocity component near the same magnitude and direction as the aircraft, resulting in the same target images. Triangulation of the target's position is not possible. The target's velocity also retains ambiguity along the direction of observation, as illustrated by the two dashed target positions in Fig. 3. Simulations based upon the dynamic and observation models of (27) and (28) have supported this heuristic argument. A moving vehicle cannot observe the states of another moving vehicle from monocular vision alone.

V. Novel Solution

The unobservability issues for monocular tracking of moving targets from a moving platform are addressed for ground targets. The ground target constraint is used to extrapolate a 3D measurement from the 2D image coordinates. The general target observation model is redefined as illustrated in Fig. 4, where the target is assumed to lie in a plane defined by nearby features.

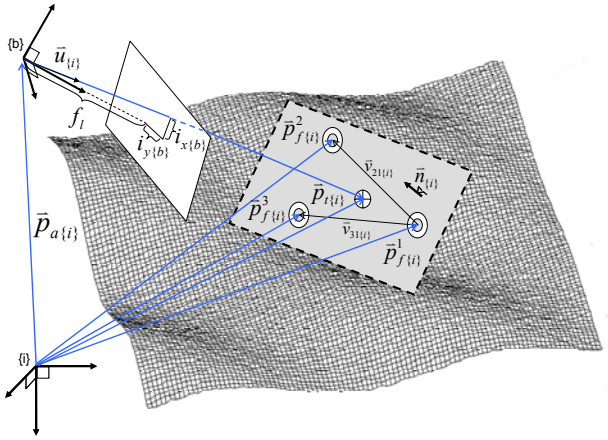

Figure 4: The relationships for the redefined target observation model.

In Fig. 4, the ground plane is defined by the three closest features to the target. The normal vector for the ground plane is given by

$$\vec{n}_{\{i\}} = \vec{v}_{21\{i\}} \times \vec{v}_{31\{i\}} \tag{29}$$

where

$$\vec{v}_{21\{i\}} = \vec{p}^{\,2}_{f\{i\}} - \vec{p}^{\,1}_{f\{i\}} \tag{30}$$

$$\vec{v}_{31\{i\}} = \vec{p}^{\,3}_{f\{i\}} - \vec{p}^{\,1}_{f\{i\}} . \tag{31}$$

The unit vector pointing from the camera toward the target is found using the image of the target and the aircraft state. This vector is expressed in the inertial frame as follows.

$$\vec{u}_{\{i\}} = \mathbf{R}_{\{ib\}} \frac{\begin{bmatrix} i_{x\{b\}} & i_{y\{b\}} & f_l \end{bmatrix}^T}{\left\| \begin{bmatrix} i_{x\{b\}} & i_{y\{b\}} & f_l \end{bmatrix} \right\|} \tag{32}$$

The position of the target, $\vec{p}_{t\{i\}}$, can now be written as a vector sum

$$\vec{p}_{t\{i\}} = \lambda \vec{u}_{\{i\}} + \vec{p}_{a\{i\}} \tag{33}$$

where λ is an unknown scalar at this point. To solve for this scalar, we use the fact that the dot product of any vector in the plane with the normal vector, $\vec{n}_{\{i\}}$, must be zero. We define a vector in the plane using the difference between a feature location and the target location.

$$\vec{n}^{\,T}_{\{i\}} (\vec{p}_{t\{i\}} - \vec{p}^{\,1}_{f\{i\}}) = 0 \tag{34}$$

Substituting (33) into (34) and rearranging results in the following.

$$\lambda \vec{n}^{\,T}_{\{i\}} \vec{u}_{\{i\}} = \vec{n}^{\,T}_{\{i\}} (\vec{p}^{\,1}_{f\{i\}} - \vec{p}_{a\{i\}}) \tag{35}$$

Solving (35) for λ gives

$$\lambda = \frac{\vec{n}^{\,T}_{\{i\}} (\vec{p}^{\,1}_{f\{i\}} - \vec{p}_{a\{i\}})}{\vec{n}^{\,T}_{\{i\}} \vec{u}_{\{i\}}} . \tag{36}$$

Substituting (36) into (33) results in an expression for the target location.

$$\vec{p}_{t\{i\}} = \left[\frac{\vec{n}^{\,T}_{\{i\}} (\vec{p}^{\,1}_{f\{i\}} - \vec{p}_{a\{i\}})}{\vec{n}^{\,T}_{\{i\}} \vec{u}_{\{i\}}} \right] \vec{u}_{\{i\}} + \vec{p}_{a\{i\}} \tag{37}$$

All of the terms on the right side of (37) are obtained from the camera sensor and estimated SLAM states. By providing a position estimate in 3D space, the observability loss for the ground target is eliminated. The measurement equation for the target, in (24), is thus modified to consist of a target position measurement with noise.

$$\vec{z}_t = \vec{p}_{tm\{i\}} = \vec{p}_{t\{i\}} + \vec{v}_{p_t\{i\}} \tag{38}$$

The covariance of the noise term, $\vec{v}_{p_t\{i\}}$, in the target position measurement can be estimated using the unscented transformation and the covariance of the feature estimates, aircraft estimate, and camera coordinates used in (29), (32), and (37).

A. Unscented Transformation

The problem of estimating the predicted mean and covariance from nonlinear functions (i.e. (29), (32), and (37)) of Gaussian random variables (GRV) has been addressed by many filtering techniques including the extended Kalman filter (EKF). The EKF assumes that the second and higher order terms of the Taylor series expansion of a nonlinear function are negligible. With this assumption, the new mean is predicted using the nonlinear function, and the new covariance is predicted using the first order term or Jacobian of the nonlinear function. While effective for many applications, highly nonlinear systems demand a higher fidelity mean and covariance estimation technique. Central to the unscented Kalman filter (UKF), the UT addresses these deficiencies by using a sampling technique that improves covariance and mean estimates.

The primary concept of the UT involves selecting a set of deterministically chosen weighted "sigma points" (vectors), which have a known mean and covariance. These points are propagated through the nonlinear model, resulting in transformed points. The statistical data represented by the transformed points approximates the true probability density function. It has been shown that the UT accurately predicts the mean and covariance to the second order and has similar computational cost as the EKF [4]. This paper does not present the details of the UT and its use in the UKF. The open literature addresses many papers on this subject, specifically [4] and [5].

Various sigma point sets exist. The set used for the UT in this paper is known as the basic symmetric set. Intuitively, the symmetric sigma points represent vectors in N_x dimensional space where N_x is the number of GRVs defining the nonlinear equations. These points are perturbed about the mean estimate by an amount that properly models the current mean and covariance. For a thorough description of this and other advanced sigma points see [4].

B. Loosely Coupled Estimation Solution

While the UT provides a method for determining a mean measurement with predicted covariance, care must be taken when applying the target observation model given by (38) within a tightly coupled Kalman filter (KF). Tightly coupled, in this paper, refers to the combination of the target states and SLAM states within a single filter. A fundamental assumption made within the derivation of the KF requires orthogonality between measurement noise terms and the estimated state covariance [10]. By using (38), which is a function of the current state estimate and camera measurement, within a tightly coupled filter, correlation will be induced. This problem can be eliminated by estimating the target states within another loosely coupled KF. Fig. 5 illustrates the high-level perspective of the loosely coupled estimators and their relationships.

The SLAM model consists of state equations (25) and (26). An UKF is utilized to estimate the aircraft and feature states due to the noise terms and nonlinear nature of the state equations. The input and output measurements are contained in dashed boxes, which represent the conversion from continuous to discrete time. The output of this SLAM UKF estimator, including the mean values for aircraft position, aircraft orientation, and feature locations and their

associated error covariance terms, becomes the input to the UT block. Another input to this block includes the output measurements of the target location in image space coordinates. The UT then converts the input means and covariance into a target position measurement and covariance matrix estimate via the relationships defined by (29), (32), and (37). The output of the UT block becomes the output measurement to the target KF. The input to the target model is simply the random walk zero mean noise term. A basic Kalman Filter is used for the target estimator since the target state equations are linear.

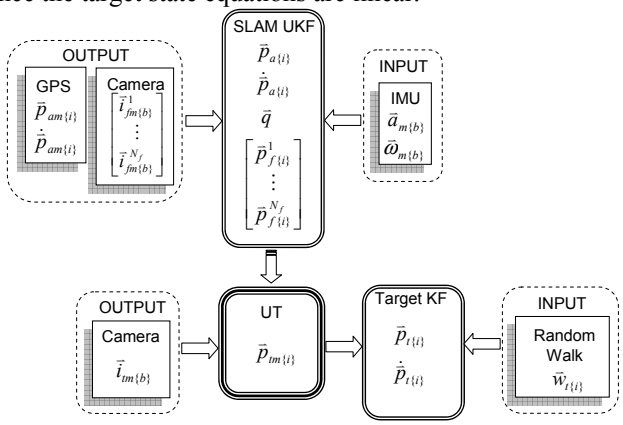

Figure 5: The relationships between the SLAM UKF, UT, and the Target KF.

VI. Simulation Results

Simulation has been used to test the validity of the SLAM and target tracking estimators discussed above. The analysis includes assessment of the convergence of feature locations, the aircraft state estimation accuracy as gained by SLAM estimation techniques over the typical GPS/INS estimation, and the convergence of the target tracker to a known value.

The simulation requires accurate modeling of an aircraft's dynamic equations, and the generation of IMU, GPS, and camera data for filter estimation. A conventional six degree of freedom aircraft model, taken from [11], is implemented in MATLAB/SIMULINK. An autopilot in the simulation allows tracking of commands from a straight-line waypoint navigator. The state outputs from the modeled aircraft, including accelerations and angular velocities, are combined with zero mean random noise in order to simulate accelerometer and gyro measurements. The aircraft position and velocity outputs are also corrupted with noise to model GPS measurements. The noise statistics were approximated from a commercial grade UAV IMU. The UKF estimator algorithm is contained within a discrete time S-Function operating at 25Hz. The UKF correction occurs at 5Hz when new simulated GPS and image measurements are available.

The environment for simulation is illustrated in Fig. 6. The aircraft begins flight in a northwest direction with a nominal velocity of 44.704m/s and altitude 100m above a ground plane, defined by features. More features are used in the simulation than shown in the illustration, although the generally layout of the features, which allows easy reference, is similar. Some environments with complex 3D structure may require additional programming logic to extract the three features that best define the local target ground plane at any given time. At 750m to the north of the aircraft's initial position, the ground terrain begins sloping downward at approximately 20 degrees. The target's initial position is located at 550m to the north of the aircraft's initial position. After 14 seconds of flight, the target, heading due north with the same nominal velocity as the aircraft, comes into view. After the target travels 200m to the north of its starting position, it turns and follows the slope maintaining its same velocity. The slope provides a proof of concept analysis emphasizing the ability of this method to track targets in undulating terrain. After environment definition, the filters are initialized with state estimates that are perturbed from their true values with amounts consistent to the initial error covariance matrices.

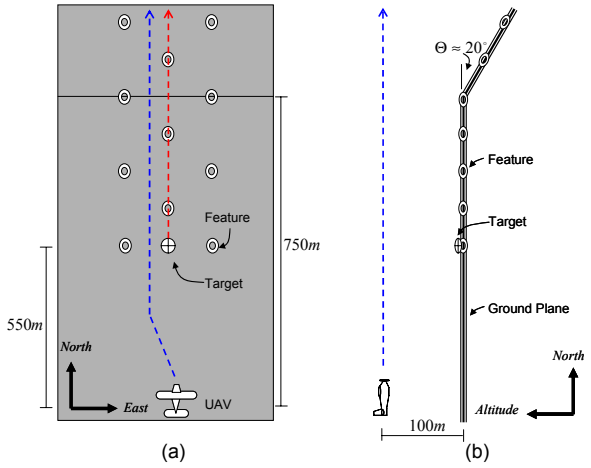

Figure 6: Simulation environment (a) top view and (b) side view.

The error magnitude of a feature's location estimate is shown in Fig. 7. During the first ten seconds of flight, the feature is not in view. Once in view, the UKF dynamically updates and improves the position estimate of the feature. Other features exhibit similar behavior.

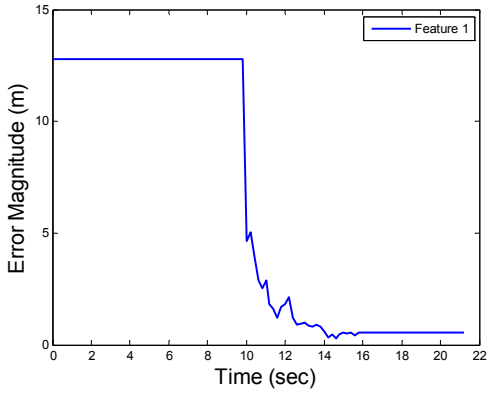

Figure 7: Feature 1 position error magnitude.

The next plot emphasizes the orientation accuracy for the aircraft gained through SLAM estimation as compared to a traditional GPS/INS estimation. As the features come into view, the SLAM filter improves the estimated aircraft

orientation, especially pitch. This behavior is expected because the observation model for the features depends heavily upon the position and orientation of the aircraft. In fact, the EKF GPS/INS solution actually developed a substantial orientation bias.

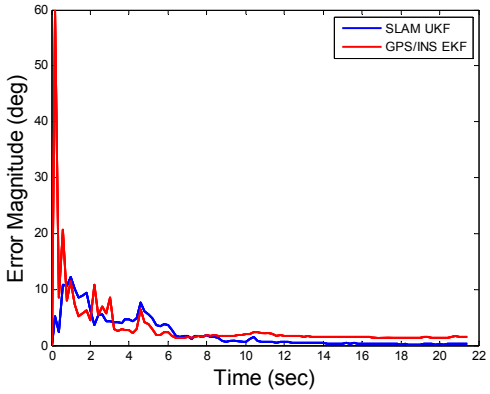

Figure 8: RMS aircraft orientation error magnitude.

The target position error illustrates the effectiveness of the target tracking technique developed. Fig. 9 plots the target position error magnitude as a function of the time during which the target is observed. The error peak between three and four seconds corresponds to the point when the target vehicle turns down the slope. While the error never converges completely to zero, the target state estimates do track the target. This technique may be used to provide tracking information for a range of applications.

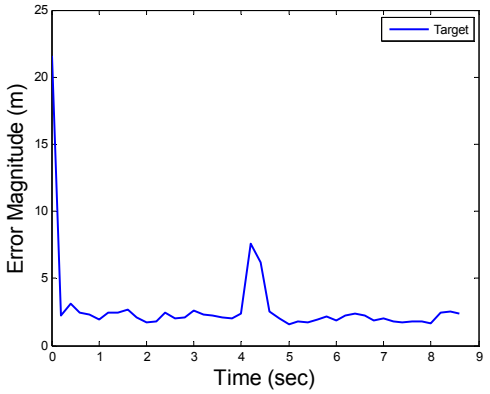

Figure 9: Target position error magnitude.

VII. Conclusion

A novel method has been presented for the tracking of ground targets observed from camera images taken from a moving platform. The observability problems caused by using a monocular camera are addressed using loosely coupled SLAM and target state estimators and by utilizing the UT to obtain measurement mean and covariance data for the target position as a function of the SLAM states and camera measurements. Simulation shows the effectiveness of the method presented. The relative ease of implementation of this method warrants its consideration for several practical applications and additional experimental research.

References

[1] E. Semerdjiev, L. Mihaylova, and X. Li, "An Adaptive IMM Estimator for Aircraft Tracking," *Proceedings of the International Conference on Information Fusion*, Sunnyvale, CA, July 1999, pp. 770-776.

[2] A. Adrien, D. Filliat, S. Doncieux, and J. Meyer, "2D Simultaneous Localization and Mapping for Micro Air Vehicles," *European Micro Air Vehicle Conference and Flight Competition*, 2006.

[3] J. Kim and S. Sukkarieh, "Airborne Simultaneous Localisation and Map Building," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Taipei, Taiwan, September 2003.

[4] S. Julier and J. Uhlmann, "Unscented Filtering and Nonlinear Estimation," *Proceedings of the IEEE*, vol. 92, no. 3. March, 2004.

[5] R. Merwe and E. Wan, "Sigma-Point Kalman Filters for Integrated Navigation," in *Proceedings of the 60th Annual Meeting of The Institute of Navigation (ION)*, Dayton, OH, June 2004.

[6] B. Stevens and F. Lewis. *Aircraft Control and Simulation*. New Jersey: John Wiley and Sons, Inc., 2003.

[7] M. Yeddanapudi, Y. Bar-Shalom, and K. Pattipati, "IMM Estimation for Multitarget-Multisensor Air Traffic Surveillance," *Proceedings of the IEEE*, vol. 85, no. 1, January 1997, pp. 80-94.

[8] M. Farmer, R. Hsu, and A Jain, "Interacting Multiple Model (IMM) Kalman Filters for Robust High Speed Human Motion Tracking," *Proceedings of the International Conference on Pattern Recognition*, vol. 2, 2002, pp. 20-23.

[9] C. Chen. *Linear System Theory and Design*. 3rd Edition. New York: Oxford University Press, 1999.

[10] F. Lewis and V. Syrmos. *Optimal Control*. 2nd Edition. New York: John Wiley & Sons, Inc., 1995.

[11] R. Nelson. *Flight Stability and Automatic Control*. New York: McGraw-Hill, Inc., 1989.

[12] A. Davison, I. Reid, N. Molton, and O. Stasse, "MonoSLAM: Real-Time Single Camera SLAM," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, 2007, pp. 1052-1067.

[13] M. Bryson and S. Sukkarieh, "Bearing-Only SLAM for an Airborne Vehicle," *Australian Conference on Robotics and Automation*, Sydney, Australia, 2005.