On the observability of indirect filtering in vehicle tracking and localization using a fixed camera

Perera L.D.L. and Elinas P.
Australian Center for Field Robotics
The University of Sydney
Sydney, NSW, Australia
{l.perera, p.elinas}@acfr.usyd.edu.au

Abstract – *In several vehicle tracking and localization applications, the initial position of a vehicle may be given by GPS measurements or other means. However, the information required for accurate tracking after initialization may only be available intermittently or not at all. In this paper, we demonstrate that the indirect or error form of state variables can be used in accurate bearing only tracking of a vehicle when the GPS measurements of its location are discontinued for some reason. Using Piece-wise Constant Systems Theory of Observability Analysis and the indirect form of the state variables for a constant velocity model we show that an object moving in a two-dimensional environment tracked by bearing only measurements using a fixed monocular camera is fully observable. We experimentally verify the theoretical results with simulations and real data from a fixed monocular camera tracking a pedestrian consistently when GPS measurements are discontinued.*

Keywords: Tracking, localization, observability.

1 Introduction

We consider the problem of localizing and tracking a moving object by fusing data from a GPS sensor and a single camera. Our motivation is from an application in open pit mining where the accurate localization of vehicles and mining personnel as they go about the pit is important for both controlling the process accurately and increasing safety.

In this paper, we show that given an initial estimate of an entity's position using GPS, we can improve our estimate of its location by integrating data from a remote camera. We also show that when the GPS signal is lost and given that the system has been initialized, we can continue localizing the entity using only a single camera until GPS tracking is restored.

We perform the necessary observability analysis to show that our method is valid. In addition, we perform an experimental evaluation using both simulated data and real data localizing a pedestrian.

2 Problem statement and previous work

It is common practice in open pit mining that vehicles are localized using GPS [9]. Accurately knowing the position and velocity of any vehicle as it goes about the mine is important for supporting automation systems. Due to the geometry of an open pit mine, there are several locations where GPS fails [7] because of an insufficient number of visible satellites, multi-path effects of local terrain and periodic signal blockage due to foliage or places having restricted view of the sky.

We consider solving the problem of continuous object localization using a combination of GPS and a monocular camera. A visual tracking system can be used to detect entities in video [15]. This information can be used to provide continuous localization in $3D$. We show that when an object's location is initially known we can accurately estimate its position by observing it from a stationary camera. Moreover, we show that when data from both sensors, GPS and camera, are available, we can get a more accurate estimate of the object's location.

Even though our motivation is based on the automation of open pit mining operations, there are several other vehicle tracking and localization applications [3] where you know the initial position of the vehicle by GPS or by other means but adequate information after initialization essential for accurate tracking is often hard to obtain. Bearing only tracking and localization of a missile from a known initial position [13] and aircraft tracking in and around airports [4] to improve traffic capacity and safety are some example applications our work also addresses.

Several sensors such as Inertial Measurement Units (IMU) [11], laser, vision and radar are often used to complement or aid GPS systems in vehicle tracking and localization using sensor fusion algorithms. A comprehensive and detailed review of the use of such multi-sensor fusion systems is available in [3]. In particular, use of computer vision in vehicle tracking and localization with GPS is demonstrated in [8, 10, 14]. Hsien-Chou et al. [10] describe a GPS-based object detection and tracking algorithm in a ubiquitous camera environment. Subong et al. [14] describe a real-time approach to calculate the three-dimensional location of a fixed target detected by a gimbaled camera in a fixed-wing experimental unmanned aerial vehicle (UAV) equipped with a single-antenna GPS receiver on board. Heimes at al. [8] describe how to combine passive GPS and map-based route guidance with model-based machine vision in order to automatically assess or even execute driving maneuvers in inner-city traffic situations.

Although all of the above approaches use machine vision as a complementary sensor system with GPS in object tracking and localization, none of the work validates the use of computer vision to complement GPS using an observability analysis which is one of the main contributions of this paper. More specifically, we show that the indirect or error form of state variables can be used in accurate bearing only tracking of an object after initialization or when the GPS measurements of its location are discontinued. Moreover, we show that the error form of the state variables of a constant velocity model of an object moving in a $2D$ environment tracked by bearing only measurements is fully observable.

The rest of this paper is structured as follows. In Section 3, we provide a brief overview of piece-wise systems theory essential for the observability analysis of Section 5.3. In Section 4, we develop the indirect or error form of $2D$ vehicle tracking and localization. We dive into more details on visual tracking in Section 5. Finally, in Section 6 we perform a comprehensive evaluation of our method using real and simulated data and conclude in Section 7.

3 Piece-wise constant systems theory

A system is said to be observable at time t_0 if its state vector at time t_0, $\mathbf{x}(t_0)$ can be determined from the measurements in t_f, $t_0 < t_f$ which is finite. Several linear and nonlinear techniques are used in observability analysis of engineering systems [1]. The piece-wise constant systems theory in particular assumes that systems are piece-wise constant over sufficiently small time segments and uses linear systems theory in the observability analysis. Use of linear systems theory in piece-wise constant systems theory provides advantages such as the possibility of using simple linear state space analysis techniques to access all state variables and simplified observer design [5, 6]. Here, we briefly summarize the piece-wise constant systems theory for continuous time systems.

Let a system be defined as follows,

$$\dot{\mathbf{x}}(t) = \mathbf{F}\mathbf{x}(t) \tag{1}$$

$$\mathbf{z}(t) = \mathbf{H}\mathbf{x}(t) \tag{2}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ represents the state variables being modelled and $\mathbf{z}(t) \in \mathbb{R}^m$ represents the measurements from the system, $\mathbf{F}$ and $\mathbf{H}$ are the process and measurement model transition matrices respectively. The observability matrix $\mathbf{O}_j$ of the system in time segment j is then defined as

$$\mathbf{z}_j(t_j) = \mathbf{O}_j x(t_1) \tag{3}$$

where $\mathbf{z}_j(t_j)$ is the concatenated vector consisting of the vector $\mathbf{z}(t_j)$ and its $n-1$ derivatives, t_1 is the initial time segment and

$$\mathbf{O}_j = \begin{bmatrix} (\mathbf{H}_j)^T & (\mathbf{H}_j\mathbf{F}_j)^T & \dots & (\mathbf{H}_j(\mathbf{F}_j)^{n-1})^T \end{bmatrix}^T \tag{4}$$

where $\mathbf{F}_j$ and $\mathbf{H}_j$ are the process and measurement model transition matrices in time segment j. The total observability matrix $\mathbf{O}(r)$ for r time segments is then defined as follows.

$$\mathbf{Z}(r) = \mathbf{O}(r)\mathbf{x}(t_1) \tag{5}$$

where

$$\mathbf{O}(r) = \begin{bmatrix} \mathbf{O}_1 \\ \mathbf{O}_2 e^{\mathbf{F}_1\Delta_1} \\ \mathbf{O}_3 e^{\mathbf{F}_2\Delta_2} e^{\mathbf{F}_1\Delta_1} \\ \dots \\ \mathbf{O}_r e^{\mathbf{F}_{r-1}\Delta_{r-1}} e^{\mathbf{F}_{r-2}\Delta_{r-2}} \dots e^{\mathbf{F}_1\Delta_1} \end{bmatrix} \tag{6}$$

and

$$\mathbf{Z}(r) = \begin{bmatrix} \mathbf{z}_1^T & \mathbf{z}_2^T & \mathbf{z}_3^T & \dots & \mathbf{z}_r^T \end{bmatrix}^T \tag{7}$$

Here, $e^{\mathbf{F}_j\Delta_j}$ terms in $\mathbf{O}(r)$ account for the transition of the state variables $\mathbf{x}(t_j)$ to that of the initial time segment $\mathbf{x}(t_1)$. The piece-wise constant systems theory states that if and only if the rank of $\mathbf{O}(r)$ at any segment r is equal to the dimension of the state vector $\mathbf{x}(t)$ then the system is fully observable. This follows from the fact that the rank of $\mathbf{O}(r)$ determines the existence of a unique solution for $\mathbf{x}(t_1)$ in (5) and from the definition of the observability. We can use the simplified observability matrix $\mathbf{O}_s(r) = \begin{bmatrix} \mathbf{O}_1^T & \dots & \mathbf{O}_r^T \end{bmatrix}^T$ in the observability analysis [5] if $Null(\mathbf{O}_j) \subset Null(\mathbf{F}_j)$ for all $1 \leq j \leq r$.

4 2D vehicle tracking and localization in indirect form

The problem of vehicle tracking and localization where the vehicle motion is not represented by a vehicle kinematic model can be formulated as follows.

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) + \eta_1(t) \tag{8}$$

$$\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t)) + \eta_2(t) \tag{9}$$

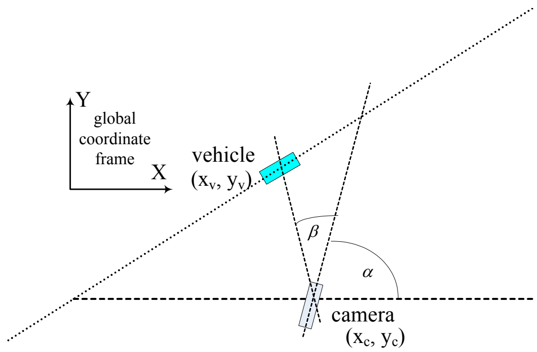


Figure 1: Camera and vehicle locations in the global coordinate frame.

where $\eta_1(t)$ and $\eta_2(t)$ are uncorrelated zero mean process and measurement noise terms with covariance $\mathbf{Q}(t)$ and $\mathbf{R}(t)$ respectively, $\mathbf{f}(.)$ is the process model and $\mathbf{h}(.)$ is the measurement model. The state vector of the vehicle is $\mathbf{x}(t) = \begin{bmatrix} x_v & v_x & y_v & v_y \end{bmatrix}^T$ where x_v, y_v, v_x and v_y are the x coordinate, y coordinate, velocity in x axis direction and velocity in y axis direction of the vehicle respectively. The indirect or error form of the vehicle tracking and localization problem excluding the noise terms from the equations for simplicity is as follows.

$$\delta\dot{\mathbf{x}}(t) = \mathbf{F}\delta\mathbf{x}(t) \tag{10}$$

$$\delta\mathbf{z}(t) = \mathbf{H}\delta\mathbf{x}(t) \tag{11}$$

where $\delta\mathbf{x}(t)$ and $\delta\mathbf{z}(t)$ are the errors between the true value and the estimated value of the state variables and the measurements respectively. The other terms of (10) and (11) are,

$$\delta\mathbf{x}(t) = \begin{bmatrix} \delta x_v & \delta v_x & \delta y_v & \delta v_y \end{bmatrix}^T \tag{12}$$

$$\mathbf{F} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \tag{13}$$

$$\mathbf{H} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \tag{14}$$

$$\delta\mathbf{x}(t) = \mathbf{x}_{true}(t) - \hat{\mathbf{x}}(t) \tag{15}$$

where $\mathbf{x}_{true}(t)$ is the true value of $\mathbf{x}(t)$ and $\hat{\mathbf{x}}(t)$ is the predicted value of $\mathbf{x}(t)$. The value of $\hat{\mathbf{x}}(t)$ is usually calculated from the estimated value, odometry or initialized value where appropriate. Hence, during a certain interval if you know the value of $\hat{\mathbf{x}}(t)$ the estimated value of $\mathbf{x}(t)$ is the sum of $\hat{\mathbf{x}}(t)$ and the estimated value of $\delta\mathbf{x}(t)$. In the following discussion we show that $\delta\mathbf{x}(t)$ is fully observable in the bearing only vehicle tracking and localization problem if we use GPS for the initial estimation and subsequent estimated value of $\mathbf{x}(t)$ in determining $\hat{\mathbf{x}}(t) + \delta\mathbf{x}(t)$.

We use a constant velocity model to represent the vehicle kinematic model. When the direct form of the 2D vehicle tracking and localization problem is considered, it is fully observable when the GPS measurements of the vehicle locations are available. Since,

$$\mathbf{f}(\mathbf{x}(t)) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{16}$$

$$\mathbf{h}(\mathbf{x}(t)) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{17}$$

$$\mathbf{O} = \begin{bmatrix} (\mathbf{H})^T & (\mathbf{HF})^T & (\mathbf{HF}^2)^T & (\mathbf{HF}^3)^T \end{bmatrix}^T \tag{18}$$

and $\mathbf{O}$ is full rank.

Consider now the indirect representation of the $2D$ vehicle tracking and localization algorithm

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{19}$$

When the GPS fix of the vehicle location is available $\mathbf{H}$ is given by,

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{20}$$

Hence the observability matrix in the first time segment

$$\mathbf{O}_1 = \begin{bmatrix} (\mathbf{H})^T & (\mathbf{HF})^T & (\mathbf{HF}^2)^T & (\mathbf{HF}^3)^T \end{bmatrix}^T \tag{21}$$

has a rank of 4. Hence, the vehicle tracking and localization system (10)-(11) of error states is observable.

5 2D vehicle tracking and localization using a monocular camera

Let there be a vehicle moving on a $2D$ horizontal plane as shown in Figure 1 with x_v and y_v as lateral and longitudinal coordinates with respect to a global coordinate frame. The velocities of the vehicle in lateral and longitudinal directions are v_x and v_y. Let the camera optical centre be at a point given by lateral and longitudinal coordinates x_c and y_c and α be the angle of the camera optical axis with the lateral direction all with respect to the selected global coordinate frame (Figure 1). It is assumed that the camera optical axis is parallel to the ground plane.

5.1 Camera calibration

The pinhole model [17] of the camera with perspective projection is used in the camera modelling and calibration. The method described in [17] and [16] was utilized in the camera calibration. This approach requires a camera to observe a planar pattern shown at a few different orientations. Either the camera or the fixed

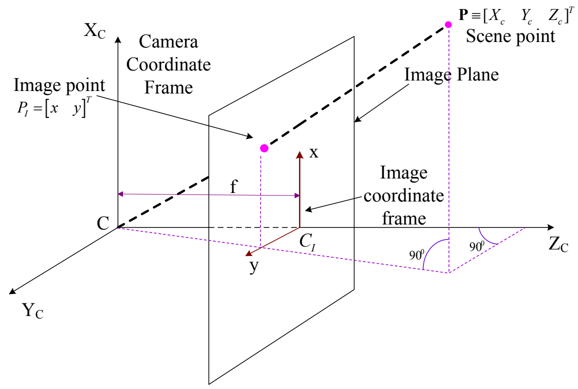


Figure 2: Camera calibration.

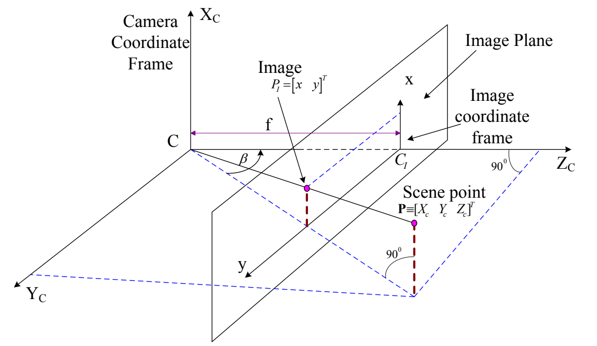


Figure 3: Bearing measurement calculation.

pattern can be moved in making observations. This method provides the calibration parameters as a closed form solution.

Let a scene point $P = \begin{bmatrix} X_c & Y_c & Z_c \end{bmatrix}^T$ be in camera coordinate frame with the focal point C as the origin and camera optical axis as the Z axis where X_c, Y_c and Z_c denote the coordinates of the scene point in the camera coordinate frame. The principal point (image centre) C_I is the origin of the image plane which is parallel to the X_c, C and Y_c plane. Let $P_I = [x \ y]^T$ be the image of the point P in the image plane as shown in the Figure 2 where x and y are its X and Y coordinates with reference to the image coordinate frame. Let $[u_0 \ v_0]^T$ be the coordinates of the image center or the principal point on the pixel frame and $[u \ v]^T$ be the pixel coordinates of the point P in pixel coordinate frame (with u and v corresponding to x and y coordinates in the image coordinate frame) S_x and S_y be the effective sizes of a pixel in the horizontal and vertical directions respectively. The following expression can then be derived for the homogeneous (projective) coordinates of the scene point and the corresponding pixel point,

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f}{S_x} & 0 & u_0 & 0 \\ 0 & \frac{f}{S_y} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c & Y_c & Z_c & 1 \end{bmatrix} \tag{22}$$

We now denote $f_x = \frac{f}{S_x}$ and $f_y = \frac{f}{S_y}$ with both the quantities having the unit of pixels. The 3×4 matrix at the right hand side of (22) is known as the Camera Calibration Matrix.

5.2 Monocular camera measurement model

Let the camera coordinate frame and the world coordinate frame be perfectly aligned so that there is no rotation between the camera and the world coordinate frames. Using the camera calibration matrix (22) and Figure 3 it follows that,

$$tan\beta = \frac{Y_C}{Z_C} = \frac{y}{f} \tag{23}$$

But, $\frac{y}{f} = \frac{(v-v_0)S_y}{f} = \frac{(v-v_0)}{\frac{f}{S_y}} = \frac{(v-v_0)}{f_y}$. Hence, it follows that,

$$tan\beta = \frac{v - v_0}{f_y} \tag{24}$$

From Figure 1 it follows that,

$$\alpha + \beta = tan^{-1}\left(\frac{y_v - y_c}{x_v - x_c}\right) \tag{25}$$

Hence,

$$tan\beta = tan(tan^{-1}\left(\frac{y_v - y_c}{x_v - x_c}\right) - \alpha) \tag{26}$$

$$\frac{v - v_0}{f_y} = tan(tan^{-1}\left(\frac{y_v - y_c}{x_v - x_c}\right) - \alpha) \tag{27}$$

Therefore, the monocular camera measurement model is;

$$v = f_y tan(tan^{-1}\left(\frac{y_v - y_c}{x_v - x_c}\right) - \alpha) + v_0 \tag{28}$$

5.3 Observability analysis

Consider now the case where GPS measurements of the vehicle location are not available and the vehicle is observed by a single monocular camera located at coordinates (x_c, y_c). Hence using (9) and (28), the monocular camera measurement model is;

$$h(\mathbf{x}(t)) = f_y tan(tan^{-1}\left(\frac{y_v - y_c}{x_v - x_c}\right) - \alpha) + v_0 \tag{29}$$

Hence using (11) it follows that;

$$\mathbf{H} = \begin{bmatrix} h_1 & 0 & h_2 & 0 \end{bmatrix} \tag{30}$$

where

$$h_1 = f_y(y_v - y_c)sec^2\theta_v/r_c^2 \tag{31}$$

$$h_2 = f_y(x_v - x_c)sec^2\theta_v/r_c^2 \tag{32}$$

$$r_c^2 = (y_v - y_c)^2 + (x_v - x_c)^2 \tag{33}$$

$$\theta_v = tan^{-1}\left(\frac{y_v - y_c}{x_v - x_c}\right) - \alpha \tag{34}$$

The observability matrix of the first time segment from (21) using $\mathbf{F}_1 = \mathbf{F}$ and using $\mathbf{H}$ from (30) is:

$$\mathbf{O}_1 = \begin{bmatrix} o_1 & 0 & o_2 & 0 \\ 0 & o_1 & 0 & o_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{35}$$

where

$$o_1 = \left(f_y(y_{v,1} - y_c)sec^2\theta_{v,1}\right)/r_{c,1}^2 \tag{36}$$

$$o_2 = \left(f_y(x_{v,1} - x_c)sec^2\theta_{v,1}\right)/r_{c,1}^2 \tag{37}$$

$x_{v,i}$, $y_{v,i}$, $x_{c,i}$ and $y_{c,i}$ represent the vehicle and camera x and y coordinates in the $i-th$ time segment respectively and

$$r_{c,i}^2 = (y_{v,i} - y_c)^2 + (x_{v,i} - x_c)^2 \tag{38}$$

$$\theta_{v,i} = tan^{-1}\left(\frac{y_{v,i} - y_c}{x_{v,i} - x_c}\right) - \alpha \tag{39}$$

Therefore, the rank of $\mathbf{O}_1$ is 2. Hence, the system of error states is not observable in the first time segment. It now follows that we can transform (5) into the following form.

$$\mathbf{T}_r\mathbf{Z}(r) = \mathbf{T}_r\mathbf{O}(r)\mathbf{M}_r\mathbf{M}_r^{-1}\mathbf{x}(t_1) \tag{40}$$

where $\mathbf{T}_r$ and $\mathbf{M}_r$ represent the matrix transformations on $\mathbf{O}(r)$. We now transform $\mathbf{T}_r\mathbf{O}(r)\mathbf{M}_r$ into the following form,

$$\mathbf{U}_r = \mathbf{T}_r\mathbf{O}(r)\mathbf{M}_r = \left[\begin{array}{c|c} \mathbf{I}_R & \mathbf{P}_R \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \tag{41}$$

where R is the rank of the $\mathbf{O}(r)$ and $\mathbf{P}_R$ is the matrix resulting from this transformation. $\mathbf{P}_R$ is an identity matrix of dimension R. Hence,

$$\mathbf{U}_R = \begin{bmatrix} 1 & 0 & \left(\frac{x_{v,1}-x_c}{y_{v,1}-y_c}\right) & 0 \\ 0 & 1 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{42}$$

Let y_O and y_U be the observable and unobservable parts of the state space. Then,

$$\mathbf{y}_O = [\mathbf{I}_R|\mathbf{P}_R]\mathbf{M}_r^{-1}\mathbf{x}(t_1) \tag{43}$$

$$\mathbf{y}_U = [\mathbf{0}|\mathbf{I}_{n-R}]\mathbf{M}_r^{-1}\mathbf{x}(t_1) \tag{44}$$

where n is the dimension of the state vector. Hence, from (43) and (44) it follows that,

$$\mathbf{y}_O = \begin{bmatrix} \delta x_v + \left(\frac{x_{v,1}-x_c}{y_{v,i}-y_c}\right)\delta y_v \\ \delta x_y + \left(\frac{x_{v,1}-x_c}{y_{v,i}-y_c}\right)\delta v_y \end{bmatrix} \tag{45}$$

$$\mathbf{y}_U = \begin{bmatrix} \delta y_v \\ \delta v_y \end{bmatrix} \tag{46}$$

It follows from (35) that the observability matrix $\mathbf{O}_1$ of the 1^{st} segment is also rank deficient by 2. Let the null vectors of the observability matrix $\mathbf{O}_1$ of the 1^{st} segment be denoted by $\mathbf{n}_{1,1}$ and $\mathbf{n}_{1,2}$

$$\mathbf{n}_{1,1} = \begin{bmatrix} \left(\frac{x_{v,1}-x_c}{y_{v,1}-y_c}\right) & 0 & 1 & 0 \end{bmatrix}^T \tag{47}$$

$$\mathbf{n}_{1,2} = \begin{bmatrix} 0 & \left(\frac{x_{v,1}-x_c}{y_{v,1}-y_c}\right) & 0 & 1 \end{bmatrix}^T \tag{48}$$

However, since $\mathbf{F}_1\mathbf{n}_{1,2} \neq \mathbf{0}$, $Null(\mathbf{O}_j) \not\subset Null(\mathbf{F}_j)$, for all j such that $1 \leq j \leq r$. Hence, we cannot use the simplified observability matrix in the segment wise observability analysis. Consider, now the Total Observability Matrix of the segments one and two.

$$\mathbf{O}(2) = \begin{bmatrix} \mathbf{O}_1 \\ \mathbf{O}_2 e^{\mathbf{F}_1\Delta t} \end{bmatrix} \tag{49}$$

$$\mathbf{O}(2) = \begin{bmatrix} o_1 & 0 & o_2 & 0 \\ 0 & o_1 & 0 & o_2 \\ \bar{o}_1 & \bar{o}_2 & \bar{o}_1 & \bar{o}_3 \\ \bar{o}_1 & \bar{o}_1 & \bar{o}_1 & \bar{o}_1 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{50}$$

$$\bar{o}_1 = \frac{f_y sec^2\theta_{v,2}}{r_{c,2}^2}\left((x_{v,2} - x_c) - (y_{v,2} - y_c)\right) \tag{51}$$

$$\bar{o}_2 = \frac{f_y sec^2\theta_{v,2}}{r_{c,2}^2}\left((x_{v,2} - x_c) - (y_{v,2} - y_c)e^{\Delta t}\right) \tag{52}$$

$$\bar{o}_3 = \frac{f_y sec^2\theta_{v,2}}{r_{c,2}^2}\left((x_{v,2} - x_c)e^{\Delta t} - (y_{v,2} - y_c)\right) \tag{53}$$

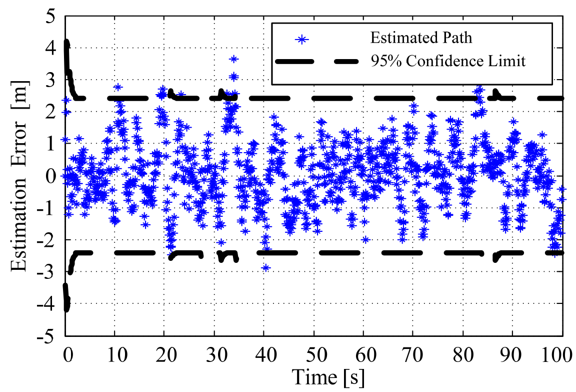
It now follows that rank of $\mathbf{O}(2)$ is equal to 4. Therefore, vehicle tracking using a single monocular camera is observable in two time segments.

6 Experimental evaluation

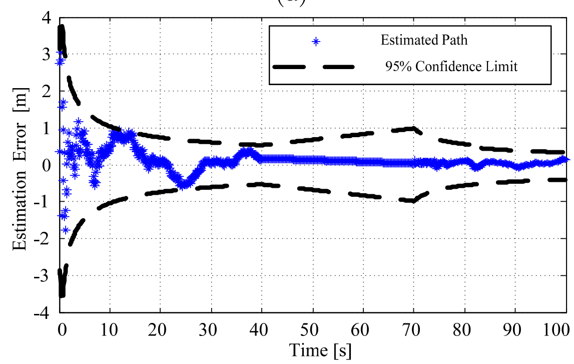
In order to verify our theoretical analysis, we performed several experiments using simulations and real data localizing a pedestrian. We present the results in the following 2 sections starting with simulation.

6.1 Simulations

For the simulation, we assume a two dimensional environment of 200×300 square meters. A vehicle is moving at constant velocities of $2ms^{-1}$ and $3ms^{-1}$ in longitudinal and lateral directions respectively subject to small acceleration perturbations. A wide field of view camera pointing perpendicular to the vehicle path is located at coordinates $(0, 300)$. We assume that the bearing measurements of the vehicle by the camera and the GPS measurements of the vehicle obtained by a GPS sensor on board the vehicle are communicated to a central location for processing. Using the information available at the central location we use constant velocity



(a)



(b)

Figure 4: Filtering errors in x direction with (a) GPS only, and (b) GPS and camera with GPS outage from 40s to 70s.

(see Equation 16) to model vehicle motion and measurements from GPS and camera to localize the vehicle using indirect filtering with the EKF. Alternatively, one can use a particle filter for tracking [2]; our observability analysis does not depend on any specific Bayesian filter implementation. For the EKF, we use zero mean Gaussian errors of Standard Deviation (STD) 3m for the GPS sensor in both longitudinal and lateral directions and a bearing error of 1^0 for the camera.

Part (a) of Figure 4 shows the localization error (shown by stars) of the vehicle in longitudinal direction with 95% confidence limits of the estimated uncertainty (shown by thick dashed lines) when only GPS is used. Part (b) of Figure 4 shows the localization error (shown by stars) in the longitudinal direction and 95% confidence limits of the estimated uncertainty (shown by thick dashed lines) when both GPS and camera measurements (bearing) are used. In this scenario we assume that GPS measurements of the vehicle's location are not available from 40s to 70s. Part (b) of Figure 4 shows that when GPS is used for initialization the filter estimate of the vehicle's location is consistent even when the GPS is discontinued for 30s.

The fact that vehicle location estimation is consistent even when the GPS outages occur verifies the observability theory. However, it is important to note that for consistent localization, the vehicle should follow the constant velocity model closely when the GPS outages occur. If the vehicle deviates significantly from the used motion model it is necessary to obtain observations from a second camera; the observability analysis of such a system is left for future work.

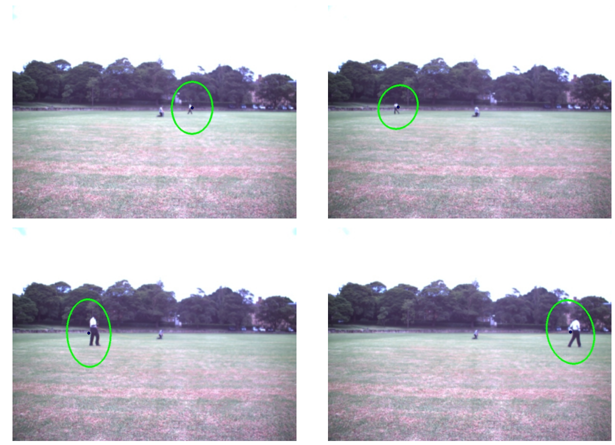


Figure 5: Examples of visual tracking using a particle filter with 100 particles. Shown are the filter mean and the 2σ covariance ellipse for a Gaussian distribution fit to the particle estimates.

6.2 Real Data

The observability theory is further verified using several pedestrian tracking experiments. Figure 5 shows examples of visually tracking a pedestrian from a fixed monocular camera. We use a Grass Hopper Gras-20S4C-C camera manufactured by Point Grey Research. The camera is located at $(-8, 0)$ pointing perpendicular to the person's path. The pedestrian is also carrying a consumer-grade GPS sensor. The camera and the GPS measurements are timestamped and available at a central location for processing. The GPS sensor has longitudinal and lateral errors of approximately 3m.

For visual tracking we used a Particle Filter [2, 12] with 100 particles; particle weights are a function of edge density within a fixed area rectangle. We fit a Gaussian distribution to the particle estimates. We show tracking examples in Figure 5 with the mean and 2σ covariance ellipse of the filter estimate.

We collected 2 data sets. For the first one, the person walked on an approximately straight line. For the second, the person walked twice in a circle of 15-meter radius. Figure 5 shows images from the second data set.

Figure 6 shows the estimated path and the 95% confidence limits of tracking the pedestrian moving along the straight line path. It is assumed that as the pedes-

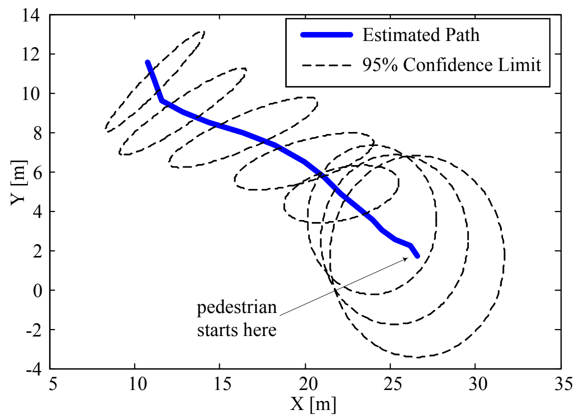
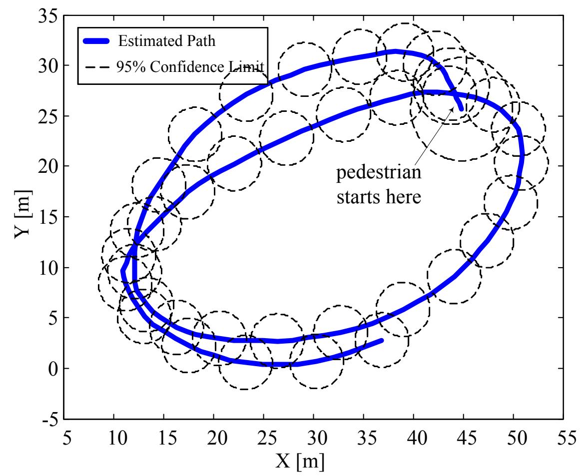


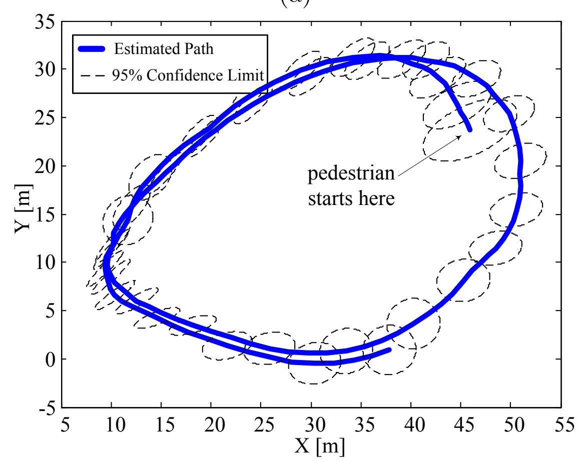
Figure 6: Tracking results of pedestrian localization (in an approximately straight line path).

trian moves half of his trajectory a GPS outage occurs until he stops at the end of his path. Diminishing area of uncertainty ellipses representing the 95% confidence bounds of the location estimates along the person's estimated trajectory in Figure 6 clearly show that even when the GPS is not available, localization is consistent thus verifying the observability analysis given in Section 5.3.

Figure 7 shows the same experimental setup with the fixed camera but the pedestrian moving along the approximately circular path in anti-clockwise direction. In this experiment, The camera is pointing towards the centre of the pedestrian's path. Part (a) of Figure 7 shows the person's estimated path (thick line) and uncertainty ellipses (dashed lines) when only GPS is used for localization. Part (b) of Figure 7 shows the same experiment when both the GPS and the camera measurements are used. Figure 8 compares the estimated uncertainties of the two cases, i.e., GPS data only and GPS plus camera data fused. Figure 8 clearly shows that the fusion of monocular camera information and GPS improves the uncertainty bound of the pedestrian's location estimate. It is also important to note that the estimated path shown in part (b) of Figure 7 more accurately reflects the true path taken by the pedestrian.



(a)



(b)

Figure 7: Tracking results fusing GPS and camera data when the pedestrian is walking in an approximately circular path.

7 Conclusions and future work

In this paper, we considered the problem of object localization and tracking fusing data from a GPS sensor and a camera. Using the piecewise constant systems theory we have shown that when the GPS signal is discontinued after some time but the object continues to be observed by a fixed monocular camera whose location is known with a certain accuracy, the error states for two dimensional point target tracking is fully observable if the object is observed for two consecutive time segments. We verified our theoretical analysis using simulations and real data tracking and localizing a pedestrian.

In future work we intend to extend the observability analysis for scenarios using multiple cameras and where GPS initialization of vehicle locations is not possible. Using more than one camera, we also intend to make the vehicle localization more robust to changing vehicle dynamics and maneuvers removing the constant velocity model assumption in the current formulation.

8 Acknowledgements

This work is supported by the Rio Tinto Centre for Mine Automation and the ARC Centre of Excellence programme funded by the Australian Research Council (ARC) and the New South Wales State Government.

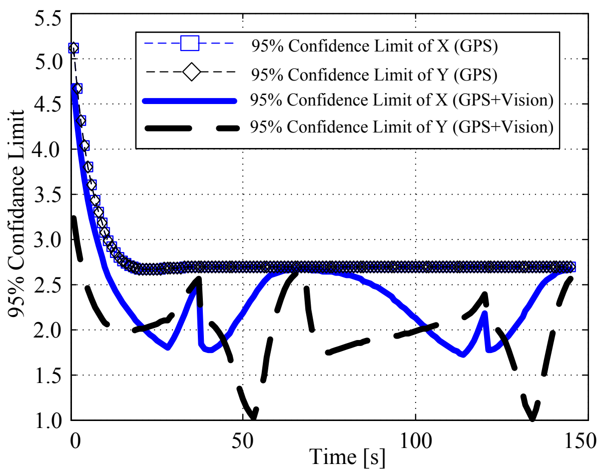


Figure 8: Confidence of tracking results fusing GPS and camera data walking in a circle.

References

[1] Y. Bar-Shalom, X.R. Li, and T. Kirubarajan, editors. *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software.* 2001.

[2] A. Doucet, Nando De Freitas, and Neil Gordon, editors. *Sequential Monte Carlo methods in practice.* 2001.

[3] H. Durrant-Whyte. A critical review of the state-of-the-art in autonomous land vehicle systems and technology. *Sandia Report SAND2001-3685*, 2001.

[4] S.C. Eng and H. Inseok. Terminal-area aircraft tracking using hybrid estimation. *Journal of Guidance, Control, and Dynamics,*, 32:836–849, 2009.

[5] D. Goshen-Meskin and I. Y. Bar-Itzhack. Observability analysis of piece-wise constant systems part I: Theory. *IEEE Transactions on Aerospace and Electronics Systems*, 28(4):1056–1068, October 1992.

[6] D. Goshen-Meskin and I. Y. Bar-Itzhack. Observability analysis of piece-wise constant systems part II: Application to inertial navigation in in-flight allignments. *IEEE Transactions on Aerospace and Electronics Systems*, 28(4):1068–1075, October 1992.

[7] M.G. Grewal, L.R. Weill, and A.P. Andrews, editors. *Global Positioning Systems, Inertial Navigation and Integration.* 2001.

[8] F.1 Heimes and H.H. Nagel. Towards active machine-vision-based driver assistance for urban areas. *International Journal of Computer Vision*, 50(1):5–34, October 2002.

[9] C. Hendricks and J. Peck. Precise positioning of blasthole drills and mining shovels using gps. Atlanta, GA, April 1996.

[10] L. Hsien-Chou and C. Pao-Tang. A novel visual tracking approach incorporating global positioning system in a ubiquitous camera environment. *Information Technology Journal*, 8(4):465–475, August 2009.

[11] G. Hyslop, D. Gerth, and J. Kraemer. Gps/ins integration on the standoff land attack missile (slam). In *IEEE Position, Location and Navigation Symposium*, pages 407–412, 1990.

[12] Michael Isard and Andrew Blake. Condensation - conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29:5–28, 1998.

[13] L. Jean and M. Riccardo. Constant-speed target tracking via bearings-only measurements. *IEEE Transactions on Aerospace and Electronic Systems*, 28(1):174–182, January 1992.

[14] S. Subong, L. Bhoram, K. Jihoon, and K. Changdon. Vision-based real-time target localization for single-antenna gps-guided uav. *IEEE Transactions on Aerospace and Electronic Systems*, 44(4):1391–1401, August 2008.

[15] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. *ACM Comput. Surv.*, 38(4):13, 2006.

[16] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *International Conference on Computer Vision (ICCV'99*, pages 666–673, Corfu, Greece, September 1999.

[17] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, October 2000.