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

Target Detection and Position Likelihood using an Aerial Image Sensor

ZuWhan Kim and Raja Sengupta

Abstract— **Sensor-based control is an emerging challenge in UAV applications. It is essential in a sensing task to account for sensor measurement errors when computing a target position estimate. Source of measurement error includes those in vehicle position and orientation measurements as well as algorithm failures such as missed detections or false detections. Incorporating such errors in aerial sensors is non-trival because of the camera's perspective geometry. This paper is about a method to incorporate such errors into target position estimates and a calibration methodology to measure the error distributions. A preliminary experiment with real flight data is presented.**

I. Introduction and Problem Statement

The importance of aerial sensing has increased due to the explosively increasing use of Geographic Information Systems (GIS) and emerging application of unmanned air vehicles (UAV's). The use of the image sensors, such as visible-range video cameras, near-infra-red cameras, and the thermal-infra-red cameras, is becoming more and more common due to their light weight and the richness of information it can provide.

There is a developing literature on vision-based UAV control and target tracking [1], [2], [3], [4], [5], [6]. The research challenges are:

- realtime target detection in the video image,
- localization of the detected target position in world or vehicle coordinates, and
- control based on the localized target position.

This paper contributes partly to the first challenge and mainly to the second one. Recent developments in computer vision have made reliable realtime object detection feasible [7]. Given the position and orientation of the vehicle and a terrain map (usually assumed to be flat), we can apply a straight forward equation to convert the target position in image coordinates to the position in world coordinates. Position derivation in this manner is straightforward when the position and orientation of the vehicle is known accurately. This is the case when the vehicle is equipped with high quality IMU's and GPS, and flown at high altitudes where six or more satellites are visible.

On the other hand, many UAV navigation systems (including ours) use relatively cheaper GPS and INS. Here, even a slight orientation error causes a significant localization error in world coordinates. In this case, we need to incorporate these IMU measurement errors in the target localizer.

This work was supported by the Office of Naval Research

ZuWhan Kim is with the Institute of Transportation Studies, University of California, Berkeley, CA 94720, USA zuwhan@berkeley.edu

Raja Sengupta is with the Department of Civil Engineering, University of California, Berkeley, CA 94720, USA sengupta@path.berkeley.edu

In addition, many errors occur in the object detection stage: missed detection and false detection. The frequency of these detection errors depends on the image resolution. For example, when the camera is oblique, the far side of the image will have a very low resolution compared to the near side. Therefore, when an object is on the far side the detection algorithm is more likely to fail than when it is on the near side. Therefore, the localization of a target should not only incorporate the measurement uncertainty of the IMU but also the varying detection errors by the object detection algorithm.

Previous efforts to generate a probability map of a target position, [3], [4], [5], have focused on the target's motion rather than the sensor measurement errors. The target detection probability (detection rate) was incorporated in some of the previous work, [3], [5], but only a constant detection rate was used.

We incorporate sensor measurement errors and varying target detection errors. Our goal is to model the target detection and position likelihoods used to build a probability map representing the location of a target. We consider two random variables: Z is a binary random variable indicating whether the target is detected on the image or not, and $\mathbf{U} = (U,V)$ is the detected target's position in image (or camera image plane) coordinates. When the target is not detected we do not have any information on $\mathbf{U}$. On the other hand, we use $\mathbf{U}$ to localize the target when it is detected. We want to know $P(\mathbf{X}|\neg Z,\hat{\mathbf{T}},\hat{\mathbf{R}})$ for the no detection case and $P(\mathbf{X}|Z,\mathbf{U},\hat{\mathbf{T}},\hat{\mathbf{R}})$ for the detection case where $\mathbf{X}$ is the true target position in the world coordinates and $\hat{\mathbf{T}}$ and $\hat{\mathbf{R}}$ are the position and orientation measurements. Applying Bayes' rule,

$$\begin{aligned} P(\mathbf{X}|\neg Z,\hat{\mathbf{T}},\hat{\mathbf{R}}) &= \alpha_Z P(\neg Z|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})P(\mathbf{X}) \text{ and} \\ P(\mathbf{X}|Z,\mathbf{U},\hat{\mathbf{T}},\hat{\mathbf{R}}) &= \alpha_U P(\mathbf{U}|Z,\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})P(\mathbf{X}), \end{aligned} \tag{1}$$

where α_Z and α_U are normalizing constants over $\mathbf{X}$, the true position of the target. $\mathbf{X}$ is independent of Z, $\hat{\mathbf{T}}$ and $\hat{\mathbf{R}}$. Therefore, we need to know $P(Z|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})$ and $P(\mathbf{U}|Z,\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})$ which are the target detection and position likelihoods, respectively.

We present the derivation of the target detection and position likelihoods in Section II. To apply the derived equation in the real application, we need to estimate the sensor measurement error. We present our error measurement methodology and experiment in Section III. Finally, we present an experimental result with real flight data in Section IV, and the conclusion in Section V.

978-1-4244-1647-9/08/$25.00 ©2008 IEEE.

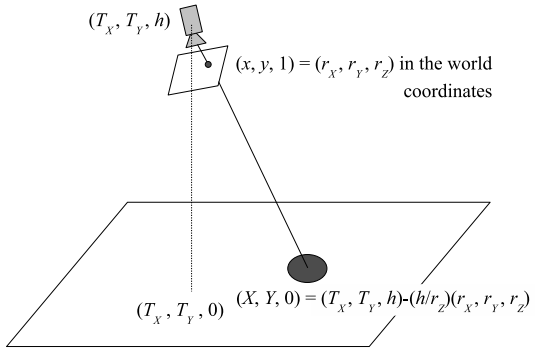


Fig. 1. Basic camera geometry when the perfect position and orientation of the camera is known.

II. Detection and Position Likelihood

A. Basic Camera Geometry

A point in image coordinates $(u,v)^T$ can be transformed to a ray (or a 3D vector), $(x,y,1)^T$, coming from the camera origin $(0,0,0)^T$ and passing through the image plane $z=1$ (see Figure 1).

The ray $(x,y,1)^T$ can be transformed (rotated and rescaled) into world coordinates, $(r_X,r_Y,r_Z)^T$, where

$$\begin{pmatrix} r_X \\ r_Y \\ r_Z \end{pmatrix} = \mathbf{R}^{-1} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \tag{2}$$

where $\mathbf{R}$ is the rotation matrix (world to camera). For convenience, we assume in the following that the camera and vehicle coordinates (positions) are the same in the world coordinate frame as adding an additional translation is straightforward.

When the vehicle's 3D position in world coordinates is $(T_X,T_Y,h)^T$ the target coordinates are

$$\begin{pmatrix} X \\ Y \\ 0 \end{pmatrix} = \begin{pmatrix} T_X \\ T_Y \\ h \end{pmatrix} - \frac{h}{r_Z} \begin{pmatrix} r_X \\ r_Y \\ r_Z \end{pmatrix}, \tag{3}$$

assuming that the target is on flat ground ($Z=0$).

B. Position Likelihood

When a target is detected at image coordinates $(u,v)^T$ or the camera coordinates $(x,y,1)^T$, we use the localization likelihood function, $P(\mathbf{U}|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})$[1], to estimate the target position in world coordinates, where $\mathbf{U}=(x,y)^T$, $\mathbf{X}=(X,Y,0)^T$ is the true target position and $\hat{\mathbf{T}}=(\hat{T_X},\hat{T_Y},\hat{h})$ and $\hat{\mathbf{R}}$ are the position and the orientation measurements of the vehicle by the onboard sensors. Note that $P(\mathbf{U}|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})$ is a PDF over $\mathbf{U}$ as it is a continuous random variable. Then,

$$\begin{aligned} P(\mathbf{U}|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) =& P(\mathbf{U}|D,\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})P(D|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) \\ &+ P(\mathbf{U}|\neg D,\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})P(\neg D|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}), \end{aligned} \tag{4}$$

where D is a binary random variable indicating whether the detection is a correct detection or a false detection. $P(D|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})$ is the probability of a detection being correct (the *detection accuracy*) and $P(\neg D|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) = 1 - P(D|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})$ is the probability of a detection being a false alarm. We may assume that these probabilities are constant or apply some resolution constraints as in Section II-C. $P(\mathbf{U}|\neg D,\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})$ is a PDF of the position given a false alarm. We can assume it is uniformly distributed over the image: $P(\mathbf{U}|\neg D,\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) = 1/\text{ImageSize}$.

Now we need to estimate

$$P(\mathbf{U}|D,\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) = \int_{\mathbf{T},\mathbf{R}} P(\mathbf{U}|D,\mathbf{X},\mathbf{T},\mathbf{R})P(\mathbf{T},\mathbf{R}|\hat{\mathbf{T}},\hat{\mathbf{R}})d\mathbf{T}d\mathbf{R}, \tag{5}$$

where $\mathbf{T}$ and $\mathbf{R}$ are for the true position and orientation of the vehicle.

We approximate $P(\mathbf{U}|D,\mathbf{X},\mathbf{T},\mathbf{R}) = \delta(\mathbf{U}-\mathbf{U}_{Tr}(\mathbf{X},\mathbf{T},\mathbf{R}))$, i.e., as a 2-D delta distribution over $\mathbf{U}$ where $\mathbf{U}_{Tr}$ is the 2-D ground-to-camera transform: $\mathbf{U}_{Tr}(\mathbf{X},\mathbf{T},\mathbf{R}) = (X_C/Z_C, Y_C/Z_C)^T$, where $\mathbf{X}_C(\mathbf{R},\mathbf{T}) = (X_C,Y_C,Z_C)^T \equiv \mathbf{R}(\mathbf{X}-\mathbf{T})$.

Next we assume the measurement errors for the position, the height, and the orientation are independent to each other. Accordingly,

$$\begin{aligned} &P(\mathbf{U}|D,\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) = \\ &\int_{\mathbf{T},\mathbf{R}} \delta(x-X_C/Z_C, y-Y_C/Z_C)P(\mathbf{T}_{XY}|\hat{\mathbf{T}}_{XY})P(h|\hat{\mathbf{h}})P(\mathbf{R}|\hat{\mathbf{R}})d\mathbf{T}_{XY}dhd\mathbf{R}. \end{aligned} \tag{6}$$

When a vehicle is moving at a fast speed but use a low-temporal resolution measurement (the GPS), the position measurement error will be large towards the vehicle's heading while the lateral error will be small because the temporal error maps to an error in the direction of the heading. Therefore, it is a good idea (see next section for the experimental validation) to assume that $P(\mathbf{T}_{XY}|\hat{\mathbf{T}}_{XY}) = P(D_F)P(D_L)$ where D_F and D_L are the difference of the positions in the forward and lateral direction, respectively. In other words, $(D_F,D_L)^T$ is $\mathbf{T}_{XY} - \hat{\mathbf{T}}_{XY}$ rotated by the yaw angle. When we further assume that the rotation measurement errors for roll (ψ), pitch (θ), and yaw (ϕ), are independent to each other, it follows

$$\begin{aligned} P(\mathbf{U}|D,\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) = \int_{\mathbf{T},\mathbf{R}} &\delta(\mathbf{U}+\mathbf{g}(\mathbf{T},\mathbf{R})P(D_F)P(D_L)P(h|\hat{h}) \\ &P(\psi|\hat{\psi})P(\theta|\hat{\theta})P(\phi|\hat{\phi})d\mathbf{T}_{XY}dhd\psi d\theta d\phi, \end{aligned} \tag{7}$$

where $\mathbf{g}(\mathbf{T},\mathbf{R}) = (-X_C/Z_C, -Y_C/Z_C)^T$.

When the measurement errors are small enough, we can approximate $\mathbf{g}(\mathbf{T},\mathbf{R})$ by

$$\begin{aligned} \tilde{\mathbf{g}}(\mathbf{T},\mathbf{R}) \approx& \mathbf{g}(\hat{\mathbf{T}},\hat{\mathbf{R}}) + D_F\frac{\partial \mathbf{g}}{\partial D_F}(\hat{\mathbf{T}},\hat{\mathbf{R}}) + D_L\frac{\partial \mathbf{g}}{\partial D_L}(\hat{\mathbf{T}},\hat{\mathbf{R}}) \\ &+ (h-\hat{h})\frac{\partial \mathbf{g}}{\partial h}(\hat{\mathbf{T}},\hat{\mathbf{R}}) + (\psi-\hat{\psi})\frac{\partial \mathbf{g}}{\partial \psi}(\hat{\mathbf{T}},\hat{\mathbf{R}}) \\ &+ (\theta-\hat{\theta})\frac{\partial \mathbf{g}}{\partial \theta}(\hat{\mathbf{T}},\hat{\mathbf{R}}) + (\phi-\hat{\phi})\frac{\partial \mathbf{g}}{\partial \phi}(\hat{\mathbf{T}},\hat{\mathbf{R}}). \end{aligned} \tag{8}$$

Then, computing Equation 7 becomes a matter of applying successive 1-D convolutions on a delta function. Intuitively, when the camera is looking downwards, the pitch error will roughly work as a y-directional 1-D convolution and the roll error will be an x-directional one. In the same way, the

[1] $P(\mathbf{U}|Z,\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})$ of Equation 1 but we omit Z for simplicity.

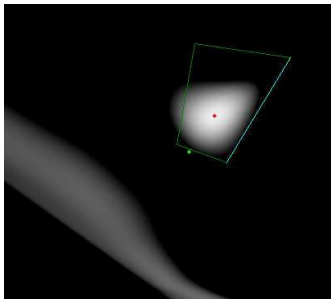


Fig. 2. A degenerated case along the intersection of the image plane and the ground plane. The probability values are exaggerated by applying a log function to clearly illustrate the problem. The quadrilateral is the camera visibility, the dot outside the quadrilateral is the estimated position of the vehicle, and the dot inside is the estimated position of the target. Such a case can be handled with a ignorable numerical error simply by thresholding the probability distribution for a large $\|\mathbf{g}\|$.

position and the height errors will also be one directional. The yaw error will not lie on a line but on an arc centered at $\mathbf{U}_{Tr}((T_X,T_Y)^T,\mathbf{T},\mathbf{R})$. However, when the angular error is small enough (usually at most several degrees) the arc can be approximated by a line.

For each of the position, height, and rotation parameters $\mathbf{g}' = -((X'_C Z_C - X_C Z'_C),(Y'_C Z_C - Y_C Z'_C))^T / Z_C^2$, where

$$\begin{array}{l}\partial\mathbf{X}_C/\partial D_F = \cos\phi(\partial\mathbf{X}_C/\partial T_X) - \sin\phi(\partial\mathbf{X}_C/\partial T_Y),\\ \partial\mathbf{X}_C/\partial D_L = \sin\phi(\partial\mathbf{X}_C/\partial T_X) + \cos\phi(\partial\mathbf{X}_C/\partial T_Y),\\ \partial\mathbf{X}_C/\partial T_X = (-R_{1,1},-R_{2,1},-R_{3,1})^T,\\ \partial\mathbf{X}_C/\partial T_Y = (-R_{1,2},-R_{2,2},-R_{3,2})^T,\\ \partial\mathbf{X}_C/\partial h = (-R_{1,3},-R_{2,3},-R_{3,3})^T,\\ \partial\mathbf{X}_C/\partial\psi = (\partial\mathbf{R}/\partial\psi)(\mathbf{X}-\mathbf{T}),\\ \partial\mathbf{X}_C/\partial\theta = (\partial\mathbf{R}/\partial\theta)(\mathbf{X}-\mathbf{T}),\ \text{and}\\ \partial\mathbf{X}_C/\partial\phi = (\partial\mathbf{R}/\partial\phi)(\mathbf{X}-\mathbf{T}).\end{array} \tag{9}$$

$\mathbf{R} = \mathbf{R}_{BC}\mathbf{R}_{WB}$ where $\mathbf{R}_{BC}$ is the body-to-camera rotation and $\mathbf{R}_{WB}$ is the world-to-body rotation. $\mathbf{R}_{BC}$ is independent of the attitude measurements, for example, $\partial\mathbf{R}/\partial\psi = \mathbf{R}_{BC}(\partial\mathbf{R}_{WB}/\partial\psi)$. Since $\mathbf{R}_{WB}$ is represented by cosines and sines of the yaw, pitch and roll angles, we can get its partial derivatives in analytic forms.

Note that the approximation in Equation 8 fails when $\|\mathbf{g}'\|$ is extremely large. Such degenerated cases happens along the intersection of the image plane and the ground plane (Figure 2). In fact, when the camera angle is very oblique, the image plane can intersect the ground plane near the vehicle. One way to work around is to set $P(\mathbf{U}|D,\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) = 0$ for a large $\|\mathbf{g}\|$. When the measurement errors are small, such thresholding will result in a small numerical error especially when $P(\neg D|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) \neq 0$.

C. Detection Likelihood

When no target is detected in image coordinates, we estimate the detection (or missed detection) likelihood: $P(Z|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})$ where Z is the binary random variable indicating whether the target is detected or not. We see that

$$P(Z|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) = \int_{\mathbf{T},\mathbf{R}} P(Z|\mathbf{X},\mathbf{T},\mathbf{R})P(\mathbf{T},\mathbf{R}|\hat{\mathbf{T}},\hat{\mathbf{R}})d\mathbf{T}dhd\mathbf{R}, \tag{10}$$

where $\mathbf{T}$, h, and $\mathbf{R}$ are for the true position, height, and orientation of the vehicle. $P(Z|\mathbf{X},\mathbf{T},\mathbf{R})$ is the detection rate. When the target is in the visible range, the detection rate depends on the resolution of the image. Otherwise, the detection rate is the false detection rate. In other words,

$$\begin{aligned}P(Z|\mathbf{X},\mathbf{T},\mathbf{R}) &= P(Z|V,\mathbf{X},\mathbf{T},\mathbf{R})P(V|\mathbf{X},\mathbf{T},\mathbf{R})\\ &\quad + P(Z|\neg V,\mathbf{X},\mathbf{T},\mathbf{R})P(\neg V|\mathbf{X},\mathbf{T},\mathbf{R})\\ &= (P(Z|V,\mathbf{X},\mathbf{T},\mathbf{R}) - P(Z|\neg V,\mathbf{X},\mathbf{T},\mathbf{R}))P(V|\mathbf{X},\mathbf{T},\mathbf{R})\\ &\quad + P(Z|\neg V,\mathbf{X},\mathbf{T},\mathbf{R}),\end{aligned} \tag{11}$$

where V is a binary random variable indicating the camera coverage and $P(Z|\neg V,\mathbf{X},\mathbf{T},\mathbf{R})$ is the false detection rate (per image frame). $V =$ true if and only if the target would appear in the image given X, T, and R based on the camera geometry.

$P(Z|V,\mathbf{X},\mathbf{T},\mathbf{R})$ depends on the image resolution: $P(Z|V,\mathbf{X},\mathbf{T},\mathbf{R}) = P(Z|V,\mathrm{res}(\mathbf{X},\mathbf{T},\mathbf{R}))$, where $\mathrm{res}(\mathbf{X},\mathbf{T},\mathbf{R})$ is the image resolution of the target at $\mathbf{X}$. When we consider a small image patch at $\mathbf{U} = \mathbf{U}_{Tr}(\mathbf{X},\mathbf{T},\mathbf{R})$ and its transform to world coordinates,

$$\begin{aligned}&\mathrm{res}(\mathbf{X},\mathbf{T},\mathbf{R}) = \mathrm{res}(\mathbf{U},\mathbf{T},\mathbf{R})\\ &= \lim_{d\to 0}\sqrt{\frac{|(\mathbf{X}_{Tr}(\mathbf{U}+\mathbf{D}_U,\mathbf{T},\mathbf{R})-\mathbf{X})\times(\mathbf{X}_{Tr}(\mathbf{U}+\mathbf{D}_V,\mathbf{T},\mathbf{R})-\mathbf{X})|}{|\mathbf{D}_U\times\mathbf{D}_V|}}\\ &= \frac{h}{r_Z}\sqrt{\frac{r_X}{r_Z}b_1+\frac{r_Y}{r_Z}b_2+b_3},\end{aligned} \tag{12}$$

where $\mathbf{X}_{Tr}$ is the image to world (ground) transform, $\mathbf{D}_U = (d,0)^T$, $\mathbf{D}_V = (0,d)^T$, $b_1 = R_{3,1}^{-1}R_{2,2}^{-1} - R_{3,2}^{-1}R_{2,1}^{-1}$, $b_2 = R_{3,2}^{-1}R_{1,1}^{-1} - R_{3,1}^{-1}R_{1,2}^{-1}$, and $b_3 = R_{1,2}^{-1}R_{2,1}^{-1} - R_{1,1}^{-1}R_{2,2}^{-1}$.

To model $P(Z|V,\mathrm{res}(\mathbf{X},\mathbf{T},\mathbf{R}))$, we can collect the image resolutions of example correct- and mis-detections. By applying a Bayesian rule:

$$P(Z|V,\mathrm{res}(\mathbf{X},\mathbf{T},\mathbf{R})) = \frac{P(\mathrm{res}(\mathbf{X},\mathbf{T},\mathbf{R})|Z,V)P(Z|V)}{P(\mathrm{res}(\mathbf{X},\mathbf{T},\mathbf{R})|Z,V)P(Z|V) + P(\mathrm{res}(\mathbf{X},\mathbf{T},\mathbf{R})|\neg Z,V)P(\neg Z|V)}, \tag{13}$$

where $P(Z|V)$ and $P(\neg Z|V)$ are assumed to be constant. Then we can fit the resolutions of the correct- and mis-detections to a pair of parametric distributions for $P(\mathrm{res}(\mathbf{X},\mathbf{T},\mathbf{R})|Z,V)$ and $P(\mathrm{res}(\mathbf{X},\mathbf{T},\mathbf{R})|\neg Z,V)$.

The camera coverage

$$P(V|\mathbf{X},\mathbf{T},\mathbf{R}) = \prod_{i\in\{\text{left, right, top, bottom}\}} s_i(\mathbf{U}_{Tr}(\mathbf{X},\mathbf{T},\mathbf{R})), \tag{14}$$

where $s_i(\mathbf{T},\mathbf{R})$ is a 2-D step function along each of the the left, right, top and bottom boundaries (in the camera coordinates).

Since we derive both $P(Z|V,\mathbf{X},\mathbf{T},\mathbf{R})$ and $P(V|\mathbf{X},\mathbf{T},\mathbf{R})$ as function of camera coordinates, in other words,

$$\begin{aligned}P(Z|V,\mathbf{X},\mathbf{T},\mathbf{R}) &= f_Z(\mathrm{res}(\mathbf{U},\mathbf{T},\mathbf{R}))\ \text{and}\\ P(V|\mathbf{X},\mathbf{T},\mathbf{R}) &= f_V(\mathbf{U},\mathbf{T},\mathbf{R}),\end{aligned} \tag{15}$$

we can apply the approximation used in Section II-B. $P(\mathbf{T},\mathbf{R}|\hat{\mathbf{T}},\hat{\mathbf{R}})$ works as a two dimensional convolution mask in the camera coordinates. Since Equation 7 is the convolution on the delta function, its result itself is the convolution mask for $P(\mathbf{T},\mathbf{R}|\hat{\mathbf{T}},\hat{\mathbf{R}})$.

D. Gaussian Approximation of the Measurement Errors

When we approximate the position and the rotation errors with Gaussian distributions or Gaussian mixtures, the target position likelihood from Equation 7 will also be a Gaussian or a Gaussian mixture distribution because the convolution of a delta function and a Gaussian function is the Gaussian function and convolutions of Gaussian functions are also Gaussian.

The target detection likelihood, $P(Z|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})$, can be obtained either numerically or analytically with some assumptions. When we store $P(Z|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})$ as a digitized matrix, we can calculate the convolution mask for each and every matrix element (in world coordinates) and apply it in the transformed camera coordinates.

If the map is too large and when we cannot afford such a computation we may consider an analytic solution:

$$P(Z|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) = f(\mathbf{U};\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) * G_{\mathbf{U}}(\mathbf{U};\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}), \quad (16)$$

where $f(\mathbf{U};\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) = (f_Z(\mathbf{U},\hat{\mathbf{T}},\hat{\mathbf{R}}) - \beta) f_V(\mathbf{U},\hat{\mathbf{T}},\hat{\mathbf{R}}) + \beta$, $\beta = P(Z|\neg V,\mathbf{X},\mathbf{T},\mathbf{R})$ is a constant, and $G_{\mathbf{U}}(\mathbf{U};\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})$ is the Gaussian (or Gaussian mixture) convolution mask for $P(\mathbf{T},\mathbf{R}|\hat{\mathbf{T}},\hat{\mathbf{R}})$. Moreover, we may use the approximation

$$\begin{aligned} &P(Z|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) = f(\mathbf{U};\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) * G_{\mathbf{U}}(\mathbf{U};\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) \\ &\approx (f_Z(\mathbf{U},\hat{\mathbf{T}},\hat{\mathbf{R}}) - \beta)(\prod s_i(\mathbf{U};\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}}) * G_{\mathbf{U}}(\mathbf{U};\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})) + \beta, \end{aligned} \quad (17)$$

The Gaussian convolution of a step function is a cumulative Gaussian distribution which does not have a closed-form expression. We approximate a one dimensional cumulative Gaussian distribution, $C(x;\mu,\sigma)$, representing a cut at the boundary with a sigmoid function (cumulative Logistic distribution): $Sigmoid(x;\mu,\sigma') = 1/(1+e^{-(x-\mu)/\sigma'})$, where the parameter σ' is given such that $Sigmoid(\mu-\sigma;\mu,\sigma') = C(\mu-\sigma;\mu,\sigma) \approx 0.159$. Then, $\sigma' \approx \sigma/\log(5.289)$.

III. Estimating Sensor Measurement Error

The estimation of detection and position likelihoods requires various parameters. First, we need to know the IMU sensor measurement errors: $P(D_F)$, $P(D_L)$, $P(h|\hat{h})$, $P(\psi|\hat{\psi})$, $P(\theta|\hat{\theta})$, and $P(\phi|\hat{\phi})$. We also need to know the detection and false detection rates: $P(D|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})$ of Equation 4, $P(Z|\neg V,\mathbf{X},\mathbf{T},\mathbf{R})$ of Equation 11, and $P(\text{res}(\mathbf{X},\mathbf{T},\mathbf{R})|Z,V)$ and $P(\text{res}(\mathbf{X},\mathbf{T},\mathbf{R})|\neg Z,V)$ of Equation 13.

A. IMU Sensor Measurement Error

The challenge of estimating IMU sensor measurement errors is that it is very difficult to obtain the ground truth. One may use a high quality gyroscope (such as a ring gyroscope) for ground truth, but such a gyroscope is difficult to purchase because of its high cost and export restrictions.

Instead, we present a vision-based calibration. We used the Piccolo II UAV autopilot by the Cloud Cap Technology Inc. for the IMU measurements. The Piccolo II combines a GPS reading with the gyroscope output to estimate the position and the orientation. Since it uses the GPS position and speed information to measure the position and the heading (for yaw) of the vehicle and also to stabilize the pitch and the roll estimates, the resulting position and orientation estimates are very inaccurate when the vehicle is on the ground or not moving fast enough. Therefore, we cannot calibrate the position and orientation of the vehicle using a calibration board.

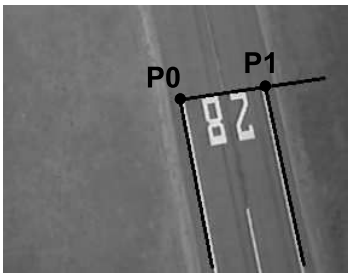


Fig. 3. A pair of long parallel lines and a perpendicular line can provide a fairly accurate position and orientation measurements of the UAV. Such a pattern is easily found in most runways. We use these vision-based measurements as ground truth to estimate the IMU measurement errors.

Instead, we flew several *calibration flights* over markings on the ground. In our past work, [8], we presented a simple calibration method to find the position and the orientation of the camera from a rectangle (or two perpendicular pairs of parallel lines), and showed that its performances compares well with calibration with grids. We use even a simpler (minimal) pattern for this experiment – a pair of parallel lines and a perpendicular line, which we can easily get from most runways (Figure 3).

The calibration procedure is similar to that of [8]. We first calculate the vanishing point $\mathbf{V}_X$ of the two parallel lines. Note that we use camera coordinates. All the points are on the image plane ($Z = 1$). The second vanishing point $\mathbf{V}_Y$ should be on a line $\overline{\mathbf{P}_0\mathbf{P}_1}$ in Figure 3: $\mathbf{V}_Y = \mathbf{P}_0 + \alpha(\mathbf{P}_1 - \mathbf{P}_0)$. Since $\mathbf{V}_X \cdot \mathbf{V}_Y = 0$,

$$\alpha = -\frac{\mathbf{V}_X \cdot \mathbf{P}_0}{\mathbf{V}_X \cdot (\mathbf{P}_1 - \mathbf{P}_0)}. \quad (18)$$

We follow [8] to estimate the rotation and the orientation from the vanishing points. The last axis $\mathbf{V}_Z = \mathbf{V}_X \times \mathbf{V}_Y$, and the rotation matrix $\mathbf{R} = (\mathbf{V}_X^T|\mathbf{V}_Y^T|\mathbf{V}_Z^T)$. The translation $\mathbf{T} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$, where

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & -p_0 \\ 0 & 1 & -q_0 \\ 1 & 0 & -p_1 \\ 0 & 1 & -q_1 \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} p_0 z_{R0} - x_{R0} \\ q_0 z_{R0} - y_{R0} \\ p_1 z_{R1} - x_{R1} \\ q_1 z_{R1} - y_{R1} \end{pmatrix}, \quad (19)$$

where $\mathbf{P}_0 = (p_0, q_0, 1)$, $\mathbf{P}_1 = (p_1, q_1, 1)$, $(x_{R0}, y_{R0}, z_{R0})^T = \mathbf{R}\mathbf{X}_0$, $(x_{R1}, y_{R1}, z_{R1})^T = \mathbf{R}\mathbf{X}_1$, and $\mathbf{X}_0$ and $\mathbf{X}_1$ are world coordinates of $\mathbf{P}_0$ and $\mathbf{P}_1$, respectively.

Based on the ground truth, we fit the IMU sensor measurement errors into parametric distributions. We use Gaussian distributions to fit the errors for this experiment but a Gaussian mixture may also be used to represent the distributions for, for example, the GPS position measurement $(P(D_F), P(D_L))$ which is supposed to be bell-shaped (shorter tails).

Note that both internal and external camera calibration is needed to convert image coordinates to the airplane's body coordinates. When using a fixed-zoom camera, the internal calibration parameters do not change and we obtain them off-line by showing a calibration grid before or after the flight. The external camera calibration (the relative rotation between the camera and the IMU) is difficult to obtain because no common reference is available between the camera and the IMU.

One way to estimate the relative rotation is to manually align the camera calibration grid to the IMU. For example, we can put a calibration board on flat ground and manually align the vehicle such that its IMU is parallel to the ground and its yaw direction is aligned to the calibration board. Then, we know the vehicle-body-to-grid rotation (a flipped identity matrix) and the grid-to-camera rotation can be obtained by a simple camera calibration procedure (for example, by finding two parallel lines and a perpendicular line – which can be automated). Then, the vehicle-body-to-camera rotation can be obtained by multiplying the two matrices.

Another way to find the rotation, which we used in this experiment, is to pick a reference image frame from the calibration flight data, where we assume that the IMU reading is correct in that particular frame. The reference image frame was manually chosen such that the resulting rotation minimizes the overall pitch, roll, and yaw errors. It can give more accurate error measurement because it does not involve any manual alignments. However, it may not pick up the bias in the measurement, for example by a wind. Therefore, the actual error can be slightly larger than the estimates.

The resulting IMU sensor measurement error estimates extracted from total 23 image frames and 3 flight paths are shown in Table I. We observe that the position error in forward direction (D_F) and the lateral direction (D_L) is very different which validates our model in the previous section.

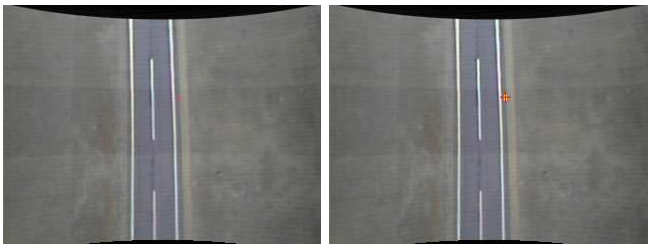
TABLE I

IMU SENSOR MEASUREMENT ERRORS

Measurement	Standard deviation of the error
Forward (D_F)	13.6m
Lateral (D_L)	2.9m
Height (h)	4.5m
Roll (ψ)	2.5°
Pitch (θ)	1.9°
Yaw (ϕ)	3.1°

B. Target Detection Error

The detection and false detection rates are obtained by counting the number of detections, missed detections and false detections from video images obtained from actual flights. We developed a user interface to manually validate correct detections and position the undetected targets. The detection accuracy (probability that a detection is a correct detection), $P(D|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})$, is assumed to be constant and it is estimated by simply counting the number of correct detections and false detections. The probability of false detection, $P(Z|\neg V,\mathbf{X},\mathbf{T},\mathbf{R})$, is also assumed to be constant and estimated by counting the number of image frames that contained the false detection.



(a) (b)

Fig. 4. A 2m × 2m red tarp was used as an experimental target and a simple Bayesian color detector was applied to detect the target: (a) an example target image and (b) the detection result.

$P(\text{res}(\mathbf{X},\mathbf{T},\mathbf{R})|Z)$ and $P(\text{res}(\mathbf{X},\mathbf{T},\mathbf{R})|\neg Z)$ require resolution parameters. The resolution is obtained by Equation 12 assuming that the position and orientation estimates are correct. We could also use the vision-based position and orientation estimates when the target is close to the calibration pattern. However, in this experiment, we used the IMU sensor results.

We used a 2m × 2m red tarp as an experimental target and applied a simple Bayesian color detector on the HLS color space to detect the target. First, the image was converted from RGB to HLS. We followed the implementation in the OpenCV library for the color space conversion. Then, a Bayesian color classifier was applied to each pixel. The detected red pixels are grouped by a connected component analysis and thresholded based on its size. An image of the target and the detection result is shown in Figure 4.

Since we have simplified the detection problem, the detection performance is very good. For example, we collected data from 392 image frames. The target was detected in 363 of them. For all detections and missed detections, the resolutions of the target were computed and used to fit $P(\text{res}(\mathbf{X},\mathbf{T},\mathbf{R})|Z)$ and $P(\text{res}(\mathbf{X},\mathbf{T},\mathbf{R})|\neg Z)$ to Gaussian distributions with the same mean. The detection accuracy, $P(D|\mathbf{X},\hat{\mathbf{T}},\hat{\mathbf{R}})$, was 0.99 and the false detection rate per frame, $P(Z|\neg V,\mathbf{X},\mathbf{T},\mathbf{R})$, was under 0.001 in our test flights.

IV. EXPERIMENTAL RESULT

For experimental evaluation, we used a fixed wing UAV equipped with a Piccolo II UAV autopilot made by the Cloud Cap Technology Inc. The true ground position of the target (Figure 4) was measured using a hand-held GPS.

An example target position likelihood is shown in Figure 5a. The likelihood is shown in world coordinates where X is pointing North and Y West. We observe that the distribution is a long ellipse shape due to the high forward-directional error (D_F). To verify if the likelihood follows the derivation in Section II, we exaggerated the error of each measurement after reducing the dominant D_F errors (Figure 5b-e). When the h-direction error is exaggerated, the distribution stretches towards the vehicle position, the

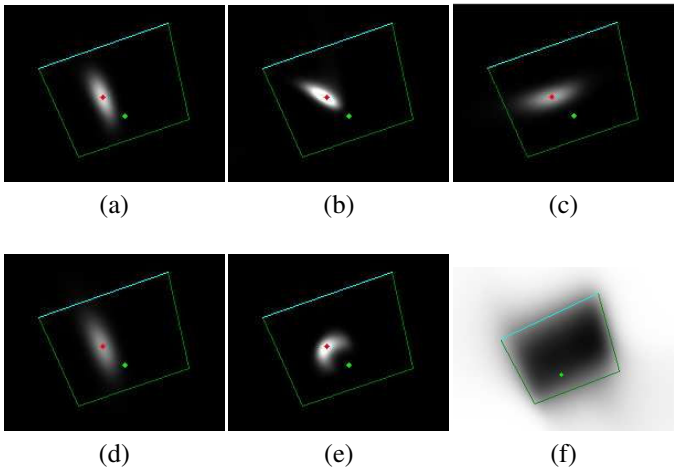


Fig. 5. (a) An example target position likelihood. The red (dark) dot is the estimated target position without applying the sensor error, the green dot is the projected position of the vehicle, and the quadrilateral boundary is the camera coverage. We see that the forward directional error (for D_F) is dominant. The rest of the figures are the likelihoods when an individual error is exaggerated: (b) h, (c) roll, (d) pitch, and (e) yaw. (f) An example target detection likelihood. Previous work only used binary camera coverage (the quadrilateral boundaries) but our estimation provides smooth boundaries.

roll error causes u-directional stretch, the pitch error causes v-directional stretch, and the yaw error causes a banana-shaped stretch. Note that the arc center of the banana-shaped stretch is supposed to be the vehicle position, but our linear approximation (Equation 8) moves its center to the mid-point between the vehicle and the target position. However, this does not introduce a significant error because such an arc curvature is observable only when the yaw error is dominant (in the example, the error was exaggerated by 10 times).

An example detection likelihood map is shown in Figure 5f. When no target is detected in the image, the target likelihood is higher outside the camera coverage than the inside. We observe that the binary camera coverage (the quadrilateral boundary) used by most previous work cannot properly model the smooth likelihood estimates obtained by our method.

The suggested derivation can be applied to various localization algorithms including Bayesian filtering and particle filtering. An example application to sensor-based path planning of a UAV is presented in [9] where the suggested derivation was used for a target search and localization.

Finally, we present a preliminary experiment to evaluate the performance of the target localization from multiple frames. When we assume that a target is not moving, we can apply a simple Bayesian reasoning assuming the independence of the individual measurements in each frame:

$$P(\mathbf{X}|\mathbf{U}_1,\hat{\mathbf{T}}_1,\hat{\mathbf{R}}_1,\mathbf{U}_2,\hat{\mathbf{T}}_2,\hat{\mathbf{R}}_2,\ldots) = \alpha P(\mathbf{X})\prod_i P(\mathbf{U}_i|\mathbf{X},\hat{\mathbf{T}}_i,\hat{\mathbf{R}}_i), \quad (20)$$

where α is the normalizing constant over $\mathbf{X}$.

We applied the above equation to a probability grid of 0.2m $\times$ 0.2m resolution, and localized the target using the maximum likelihood estimation. From two flight paths, the target was detected in 28 frames including one false detection. We compared our result with a Gaussian distribution of fixed position errors in world coordinates, where the localization result is simply the average of the estimates of individual frames. The plain Gaussian distribution resulted in a large error (22.0m) when the false detection was included and a relatively smaller error (15.0m) without counting the false detection. The false detection did not affect the localization performance of our method because it was modeled in the derivation. The localization error using our distribution was 11.2m.

V. Conclusion and Future Work

We introduced a method to incorporate sensor measurement errors when using a vision sensor to produce the target position estimates. We also presented a calibration methodology to measure the error distributions of the sensors. Our method models the orientation and position errors of an IMU and the varying detection rates corresponding to varying image resolutions. It also models the false detections. The approach can be applied to a wide range of vision- (including IR) based target localization and control tasks emerging in UAV applications. The future work is to thoroughly evaluate our model and the calibration methodology using a larger number of empirical and synthetic dataset.

VI. Acknowledgements

This work is a part of a larger UAV collaboration project under the management of Professors Raja Sengupta and Karl Hedrick, and many students of the research group contributed significantly to the experimental work. The original motivation for this work came from a group discussion with John Tisdale, Allison Ryan and David Törnqvist. The flight experiment is due to Xiao Xiao, Stephen Jackson, Mark Godwin, Nick Barton, Trevor Edmonds, and David Wood.

References

[1] E. Frew, T. McGee, Z. Kim, X. Xiao, S. Jackson, M. Morimoto, S. Rathinam, J. Padial, and R. Sengupta, "Vision-based road following using a small automonous aircraft," in *Proc. IEEE Aerospace Conference*, 2004.

[2] T. G. McGee, R. Sengupta, and K. Hedrick, "Obstacle detection for small autonomous aircraft using sky segmentation," in *Proc. IEEE Intl. Conf. on Robotics and Automation*, 2005.

[3] L. Schenato, S. Oh, S. Sastry, and P. Bose, "Swarm coordination for pursuit evasion games using sensor networks," in *Proc. IEEE Conference on Robotics and Automation*, 2005, pp. 2493–2498.

[4] Y. Yang, A. A. Minai, and M. M. Polycarpou, "Evidential map-building approaches for cooperative uav search," in *Proc. 2005 American Control Conference*, 2005, pp. 116–121.

[5] L. F. Bertuccelli and J. P. How, "Search for dynamic targets with uncertain probability maps," in *Proc. 2006 American Control Conference*, 2006, pp. 737–742.

[6] S. Rathinam, P. P. de Almeida, Z. Kim, S. Jackson, A. Tinka, W. Grossman, and R. Sengupta, "Autonomous searching and tracking of a river using an uav," in *Proc. American Control Conference*, 2007, pp. 359–364.

[7] P. Viola and M. Jones, "Rapid real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004.

[8] Z. Kim, "Geometry of vanishing points and its application to external calibration and realtime pose estimation," Institute of Transportation Studies, Research Report UCB-ITS-RR-2006-5, 2006.

[9] J. Tisdale, A. Ryan, Z. Kim, D. Törnqvist, and K. Hedrick, "A multiple uav system for vision-based search and localization," in *American Control Conference*, 2008.