# Aerophotogrammetry using UAVs: a stochastic sensor fusion approach

**Roberto Ferraz de Campos Filho, roberto.ferraz@gmail.com**
**Newton Maruyama, maruyama@usp.br**
Escola Politécnica da Universidade de São Paulo

***Abstract.*** *Photogrammetric maps are of extreme importance in order to monitor large areas periodically, for example, possible tasks might be the monitoring of forests, invasive plants, urban growth, etc. These maps are commonly built using images from satellites or planes. In order to obtain a map with real proportions, an operation of distortion of these images is realized using information provided by Ground Control Points and triangulating natural features in the scene or using another a priori known map. The utilization of an Unmanned Aerial Vehicle (UAV) provides a safer solution when compared to a plane mainly due to the non requirement of a crew. It is also a more flexible solution when compared to satellites because an UAV can fly again some hours or even minutes after the first time, while a satellite is available in some days for the same area. Some parts of the map might not be visible because of clouds and the UAV needs to fly again to recover these parts (flying below the clouds if necessary). A stochastic sensor fusion method is proposed that combines computational vision techniques, inertial sensors and GPS in order to estimate both the three dimensional sparse map and the plane position using the technique known as SLAM (Simultaneous Localization and Mapping). The complete map is generated while projecting the images into the sparse map. The main advantage of this method is that the map is constructed without the use of a priori knowledge of the terrain. The main contributions of this work are: the integration of SLAM techniques into the Aerophotogrammetry field and the development of a method that can realize a 3D mapping without the use of a priori knowledge of the terrain.*

***Keywords:*** *sensor fusion, computer vision, inertial navigation, unmanned aerial vehicle*

## 1. INTRODUCTION

Aerial photos show a lot of information that can be used for urban planning, detect invasive plants in farms, illegal deforestation, tragedies caused by storms, hurricanes, etc (Medlin *et al.*, 2000; van Klinken *et al.*, 2007). These photos cannot be used singly because distance commonly needs to be measured and the terrain generates distortions in these photos that make it impossible to do with high precision and sometimes the distance is bigger than one photo area. The Photogrammetric field makes it possible. Photos are taken by a plane or satellite and they are distorted in order to compensate the terrain and mosaicked. This mosaic needs to be georeferenced to be possible measure distances.

Each image needs to have some points measured with a GPS on the spot (ground control points), because this inform an initial orientation and compute the camera position. After that, natural features are identified in each image and they are triangulated for the purpose of computing its elevation. Using this elevation, each image is projected perspectively on the surface and the final image is formed projecting they orthogonally. This is a manual process and spends a lot of time.

This paper proposes the use of an UAV (Unmanned Aerial Vehicle) to take these pictures and also a stochastic sensor fusion approach to construct these mosaics. Simulations are made in order to validate the proposal method and is intended to be implemented it using real data collected from the UAV Apoena from XMobots (Figure 1). The advantages of using an UAV is that it is more secure than a conventional plane (there is no crew, so no one physically involved in an accident), the availability is bigger than a satellite, can operate for hours or even all day long without stress of the pilot (they can be replaced during the operation). The stochastic sensor fusion approach proposed in this paper is a technique studied in robotic field called SLAM (Simultaneous Localization and Mapping) (Durrant-Whyte and Bailey, 2006a,b). This approach is commonly used in terrain robots and there are some works using it with UAVs (Bryson *et al.*, 2009; Bryson and Sukkarieh, 2009; Törnqvist *et al.*, 2009).

The chapter 2 details the SLAM algorithm explaining the prediction model in chapter 2.1, the sensors models in chapters 2.2, 2.3 and 2.4. Chapter 3 shows some results obtained by simulation and chapter 4 concludes the paper.

## 2. METHOD

The aerophotogrammetry deals with the problem of trajectory estimation and terrain estimation separately. It needs ground control points with high precision in order to compute the camera pose since the precision of sensors embedded are not enough to be used. The technique SLAM uses sensor information to estimate the environment and simultaneously the position of the vehicle in the environment. Estimations become better because these two kinds of information (terrain and vehicle position) are statistically correlated, thus ground control points can be measured with less precision or even be discarded.

SLAM is based on stochastic filters as Kalman Filter (Kalman, 1960), Extended Kalman Filter (Jazwinski, 1970), Particle Filter (Gordon *et al.*, 1993; Schön, 2006), Marginalized Particle Filter (Schön *et al.*, 2006), etc. The Extended Kalman Filter is the chosen one to develop this application, because the models are nonlinear and the computational cost

Figure 1. UAV Apoena from XMobots

is less than Particle Filter and Marginalized Particle Filter.

These filters deal with the problem of recursive estimation of probability density function $p(x_t|y_t)$, where $x_t$ and $y_t$ are state and measurements in time $t$, respectively. In order to solve this, $x_t$ and $y_t$ are written as:

$$x_{t+1} = f(x_t, u_t, v_t) \tag{1a}$$
$$y_t = h(x_t, e_t) \tag{1b}$$

, where $v_t$ and $e_t$ are process noise and measurement noise, respectively, $f$ is the prediction model and $h$ sensor model.

This paper develops a prediction and sensor model to estimate the terrain and UAV position. The used sensors are a camera to identify natural features on the terrain, a GPS to be possible the georeferencing of the terrain and a IMU to increase the precision of the UAV pose estimation. This approach uses natural control points and does not need measurements on the spot. The deduction of the filter is not the scope of this work and a introduction of it can be found in (Welch and Bishop, 2001).

### 2.1 Prediction Model

As there are three sensors used in this work, it is necessary define some coordinate systems:

- **Earth** ($e$): The trajectory of the UAV and the terrain are estimated in this coordinate system. It is fixed to Earth.
- **Camera** ($c$): Coordinate system attached to the camera. Features are identified in this coordinate system.
- **Body** ($b$): The measurements of the IMU are made in this coordinate system.

The prediction model is based on Hol (2008) added feature coordinates $m_t$ to the state vector:

$$x_t = (b_t^e \quad \dot{b}_t^e \quad \ddot{b}_t^e \quad q_t^{be} \quad \omega_{eb,t}^b \quad \delta_{\omega,t}^b \quad \delta_{a,t}^b \quad m_{1,t} \quad \dots \quad m_{M,t})^T \tag{2}$$

, where $b_t^e$ is the position of IMU expressed in the Earth system, $\dot{b}_t^e$ is its velocity and $\ddot{b}_t^e$ its acceleration, $q_t^{be}$ its orientation with respect to the earth system, $\omega_{eb,t}^b$ its angular velocity, $\delta_{\omega,t}^b$ gyroscope bias, $\delta_{a,t}^b$ accelerometer bias and $m_{i,t}$ is the 3D position of natural features on the terrain with $i = 1 \dots M$.

Defined the state vector and the coordinates system, the prediction model is given by:

$$b_{t+1}^e = b_t^e + T\dot{b}_t^e + \frac{T^2}{2}\ddot{b}_t^e \tag{3a}$$
$$\dot{b}_{t+1}^e = \dot{b}_t^e + T\ddot{b}_t^e \tag{3b}$$
$$\ddot{b}_{t+1}^e = \ddot{b}_t^e + v_{\ddot{b},t}^e \tag{3c}$$
$$q_{t+1}^{be} = exp(-\frac{T}{2}\omega_{eb,t}^b) \odot q_t^{be} \tag{3d}$$
$$\omega_{eb,t+1}^b = \omega_{eb,t}^b + v_{\omega,t}^b \tag{3e}$$
$$\delta_{\omega,t+1}^b = \delta_{\omega,t}^b + v_{\delta_\omega,t}^b \tag{3f}$$
$$\delta_{a,t+1}^b = \delta_{a,t}^b + v_{\delta_a,t}^b \tag{3g}$$
$$m_{j,t+1} = m_{j,t} \tag{3h}$$

, where quaternion multiplication and exponential are defined as:

$$\begin{pmatrix} p_0 \\ \mathbf{p} \end{pmatrix} \odot \begin{pmatrix} q_0 \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} p_0 q_0 - \mathbf{p} \cdot \mathbf{q} \\ p_0 \mathbf{q} + q_0 \mathbf{p} + \mathbf{p} \times \mathbf{q} \end{pmatrix} \tag{4a}$$

$$exp(v) = \begin{pmatrix} \cos \|v\| \\ \frac{v}{\|v\|} \sin \|v\| \end{pmatrix} \tag{4b}$$

A model of constant acceleration is used to describe the position, velocity and acceleration dynamic. The process noise added in acceleration term is assumed to be enough to model the real movement. The quaternion dynamic is described as a constant angular velocity with noise added in the last one. Bias terms are modeled as a random walk. Finally, features are described as stationary.

### 2.2 Inertial Update Model

The IMU measurement vector is:

$$y_{IMU,t} = (y_{a,t}^T \quad y_{\omega,t}^T)^T \tag{5}$$

, where $y_{a,t}$ is the acceleration measured by the accelerometers and $y_{\omega,t}$ is the angular velocity measured by gyroscopes. Both measurements are made with respect to the body coordinate system.

The IMU sensor model is given by:

$$y_{a,t} = R_t^{be}(\ddot{b}_t^e - g^e) + \delta_{a,t}^b + e_{a,t}^b \tag{6a}$$

$$y_{\omega,t} = \omega_{eb,t}^b + \delta_{\omega,t}^b + e_{\omega,t}^b \tag{6b}$$

, where $R_t^{be}$ is the rotation from Earth coordinate system to IMU body in matricial form computed from quaternion $q_t^{be}$, $g^e$ is the gravity acceleration with respect to Earth coordinate system and $e_{a,t}^b$ and $e_{\omega,t}^b$ are measurement noises from accelerometers and gyroscopes, respectively.

### 2.3 GPS Update Model

A GPS receiver makes observations of the vehicle position with respect to the Earth coordinate system. The measured values are assumed to be metric values, *i.e.* positions in WGS84 are converted to metric values.

$$y_{GPS,t} = b_t^e \tag{7}$$

The distance between IMU and GPS is not considered in the equation because a common GPS noise (order of 10 meters) is much bigger than this distance.

### 2.4 Vision Update Model

The process of image formation consists basically in three steps: The first consists of the light incidence on the CCD sensor described as a perspective projection where the camera is modeled through a normalized pinhole camera model. The second is the distortion caused by lens. The last one is the digitalization of the image made by CCD sensor. These steps are described as:

$$p^i = (A \circ D \circ P_n)(p^c) \tag{8}$$

, where $P_n$ is the normalized pinhole camera model, $D$ is the lens distortion model, and $A$ is the digitalization model.

The normalized pinhole camera model (Hartley and Zisserman, 2000; Karlstroem, 2007; Hol, 2008; Criminisi *et al.*, 1997) represents the projection of a 3D point in the image plane. Figure 2 describes a pinhole camera model and the normalized one is obtained using $f = 1$:

$$Z \begin{pmatrix} x_n \\ y_n \\ 1 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{9}$$

Lens distortion is a phenomenon that all lens causes in images and it is generally a radial distortion that can be modeled as (Hol, 2008):

$$p_d^i = (1 + k_1 \left\|p_n^i\right\|^2 + k_2 \left\|p_n^i\right\|^4) p_n^i \tag{10}$$

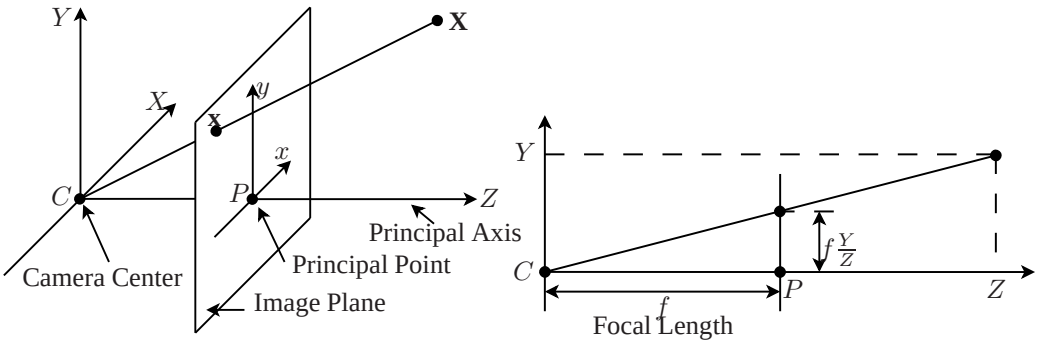


Figure 2. Pinhole camera model

, where $k_i$ are the distortion coefficients, $p_d^i$ is a point in distorted image and $p_n^i$ in normalized image.

In order to decrease the equation complexity, each image or feature is preprocessed rearranging the Equation 8 and the normalized images are used in measurement model.

$$p_n^i = (D^{-1} \circ A^{-1})(p^i) \tag{11}$$

The camera sensor model for each feature in an image is based in the one developed by Törnqvist *et al.* (2009):

$$y_{v,t} = y_{m_j,t} = \frac{1}{z_t^c} \begin{pmatrix} x_t^c \\ y_t^c \end{pmatrix} + e_{v,t}^c \tag{12}$$

, where

$$m_{j,t}^c = \begin{pmatrix} x_t^c \\ y_t^c \\ z_t^c \end{pmatrix} = R(q^{cb})R(q_t^{be})(m_{j,t} - b_t^e) + r^c \tag{13}$$

and $r^c$ is the distance between IMU and camera in the camera coordinate frame.

## 3. SIMULATIONS

Three situations of sensor fusion are compared in this section. The first one is a fusion of GPS and IMU without use of SLAM, because there is not a sensor that captures information of the environment. The second one is a fusion of IMU and Camera. This is a studied approach in situations of war scenarios when GPS may not be available or GPS failures (Bryson and Sukkarieh, 2008; Törnqvist *et al.*, 2009). The last one is the GPS, IMU and Camera fusion proposed in this work.

Data is generated with the UAV making a maneuver at 800 meters of altitude and an eight-shaped trajectory is executed by it. Features are randomly generated over a 3D surface that describes the terrain and a descriptor is associated for each one. These descriptors simulate the one that would be obtained with real images using algorithms like SIFT (Scale-invariant feature transform) or SURF (Speeded Up Robust Features), for example. They are used to match features identified in a new image with the others in the state vector. Table 1 shows the characteristics of each sensor and Table 2 the noise values used in the filter. Sensor noise values used in this work are typical of the cheapest sensors available in the market.

| | |
|---|---|
| IMU | |
| Sample rate | $100Hz$ |
| GPS | |
| Sample rate | $12.5Hz$ |
| Camera | |
| Sample rate | $12.5Hz$ |
| Resolution | 640 x 480 $pixel$ |
| Field of view | $60degrees$ |

Table 1. Sensor characteristics

Figure 3 shows the three simulations. The IMU and Camera sensor fusion are not enough to estimate the terrain with high precision because there is not a real absolute reference. Although features positions are static, it is not a real absolute

| | |
|---|---|
| Accelerometer measurement noise | $3.9m/s^2$ |
| Gyroscope measurement noise | $0.38rad/s$ |
| GPS measurement noise | $10m$ |
| Camera measurement noise | $0.035$ |
| Acceleration process noise | $0.1m/s^2$ |
| Angular velocity process noise | $0.03rad/s$ |
| Gyroscope bias process noise | $0.5mrad/s$ |
| Accelerometer bias process noise | $0.5m/s^2$ |

Table 2. Filter Settings

reference because they are estimated based in the others around it and a drift will always be present. This drift occurs when new features are being added in the filter. After that, the displacement in terrain and trajectory will be practically constant. This drift in terrain and trajectory can be noticed in figure 3(b) and figure 4.

When a GPS is fused in the filter, this drift is not present anymore because now it is the real absolute reference. The trajectory estimated by GPS and IMU configuration presents a error of $\sigma_{GPS,IMU} = 6.9m$ and the one estimated by GPS, IMU and Camera presents a error of $\sigma_{GPS,IMU,Cam} = 2.2m$. This can be seen throught Figure 4.

A feature position error analysis is made in Figure 5. The histogram of Figure 5(b) show the error caused by the drift in IMU and Camera sensor fusion configuration. The GPS, IMU and Camera configuration presents a more precise result and a error of $\sigma_{feature} = 6.6m$, that is practically a common GPS precision.

## 4. CONCLUSION

This work has been developed to generate terrain estimation with a sensor fusion approach using IMU, GPS and a camera. A SLAM approach is suggested in order to simplify the usual procedures of Aerophotogrammetry and decrease or even eliminate needing of ground control points collection. Simulations have been shown that it is possible obtaining terrain estimation with equivalent precisions of a GPS using low cost sensors. The use of an IMU makes the system robust of GPS failures once the estimation can follow some time with short drifts.

## 5. ACKNOWLEDGEMENTS

Authors would like to acknowledge CAPES for the scholarship support of the first author.

## 6. REFERENCES

Bryson, M.T., Johnson-Roberson, M. and Sukkarieh, S., 2009. "Airborne smoothing and mapping using vision and inertial sensors". *Proceedings of the 2009 IEEE International Conference on Robotics & Automation (ICRA 2009)*.

Bryson, M.T. and Sukkarieh, S., 2008. "Observability analysis and active control for airborne slam". *IEEE transactions on aerospace and electronic systems*, Vol. 44, pp. 261–280.

Bryson, M.T. and Sukkarieh, S., 2009. "Architectures for cooperative airborne simultaneous localisation and mapping". *Journal of Intelligent and Robotic Systems, Special Issue on Airborne SLAM, Springer*, Vol. 55.

Criminisi, A., Reid, I. and Zisserman, A., 1997. "A plane measuring device". *In Proc. BMVC*.

Durrant-Whyte, H. and Bailey, T., 2006a. "Simultaneous localisation and mapping (SLAM): Part I the essential algorithms". *IEEE ROBOTICS AND AUTOMATION MAGAZINE*, Vol. 13, No. 2.

Durrant-Whyte, H. and Bailey, T., 2006b. "Simultaneous localisation and mapping (SLAM): Part II state of the art". *IEEE ROBOTICS AND AUTOMATION MAGAZINE*, Vol. 13, No. 3.

Gordon, N.J., Salmond, D.J. and Smith, A.F.M., 1993. "Novel approach to nonlinear and non-gaussian bayesian state estimation". *Radar and Signal Processing, IEE Proceedings F*, Vol. 140, pp. 107–113.

Hartley, R. and Zisserman, A., 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Hol, J.D., 2008. *Pose Estimation and Calibration Algorithms for Vision and Inertial Sensors*. Master's thesis, Linköping University.

Jazwinski, A.H., 1970. *Stochastic Processes and Filtering Theory*. Academic Press.

Kalman, R.E., 1960. "A new approach to linear filtering and prediction problems". *Transactions of the ASME–Journal of Basic Engineering*, Vol. 82, No. Series D, pp. 35–45.

Karlstroem, A., 2007. *Estimação de Posição e Quantificação de Erro Utilizando Geometria Epipolar entre Imagens*. Master's thesis, Escola Politécnica da Universidade de São Paulo (in Portuguese).

Medlin, C.R., Shaw, D.R., Gerard, P.D. and LaMastus, F.E., 2000. "Using remote sensing to detect weed infestations in glycine max". *Weed Science*, Vol. 48, No. 3, pp. 393–398.

Schön, T.B., 2006. *Estimation of Nonlinear Dynamic Systems - Theory and Applications*. Ph.D. thesis, Linköping

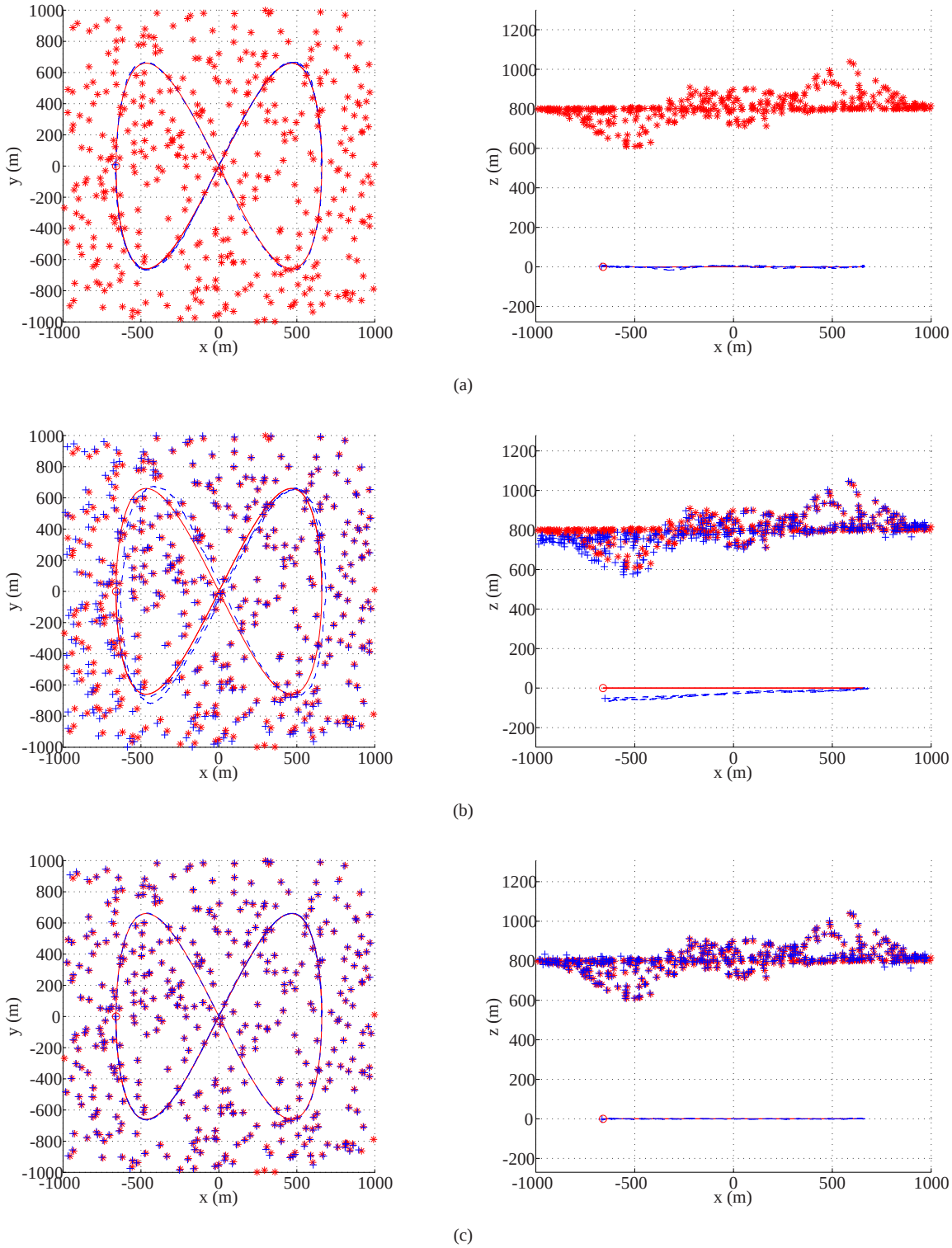



Figure 3. Terrain and trajectory estimation where red points and lines are the real features and trajectories and blue points and lines are the estimated ones. Figure (a) shows the case of GPS and IMU sensor fusion without estimation of terrain, Figure (b) IMU and Camera sensor fusion and Figure (c) GPS, IMU and camera sensor fusion

University, Linköping, Sweden.

Schön, T.B., Karlsson, R. and Gustafsson, F., 2006. "The marginalized particle filter in practice". *Proceedings 2006 IEEE Aerospace Conference*.

Törnqvist, D., Schön, T., Karlsson, R. and Gustafsson, F., 2009. "Particle filter slam with high dimensional vehicle model". *Journal of Intelligent and Robotic Systems*.

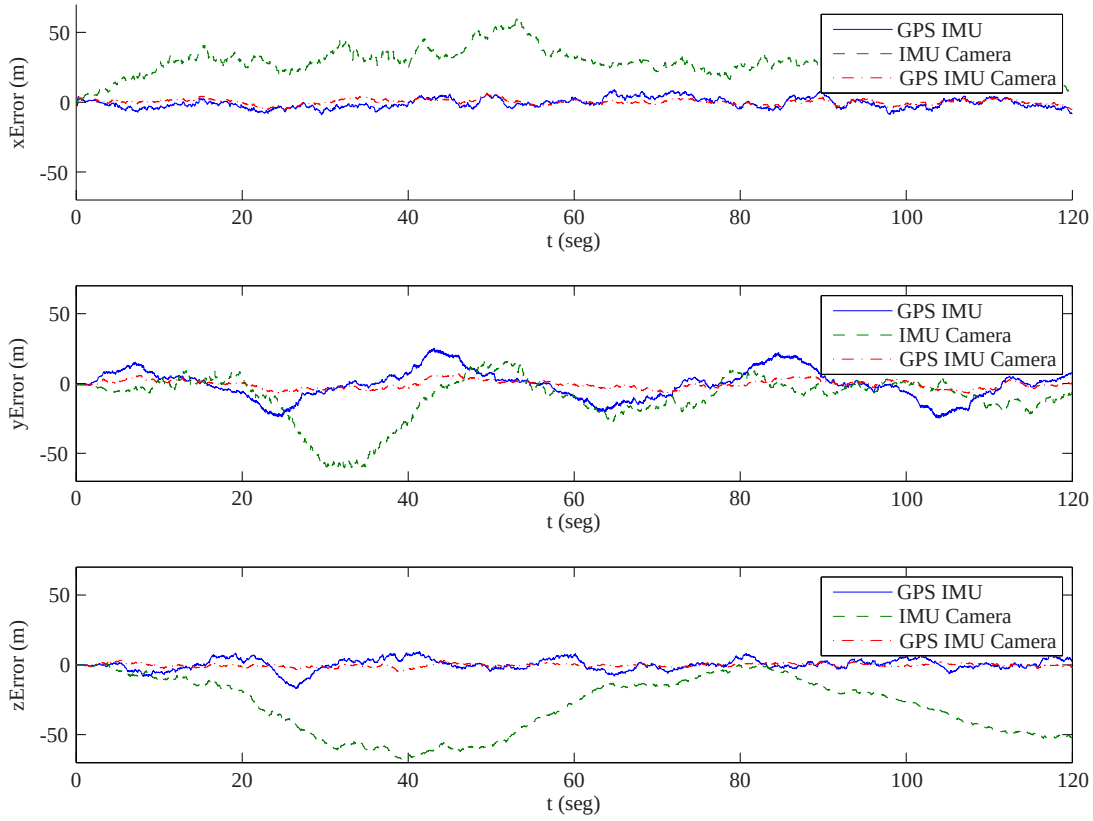
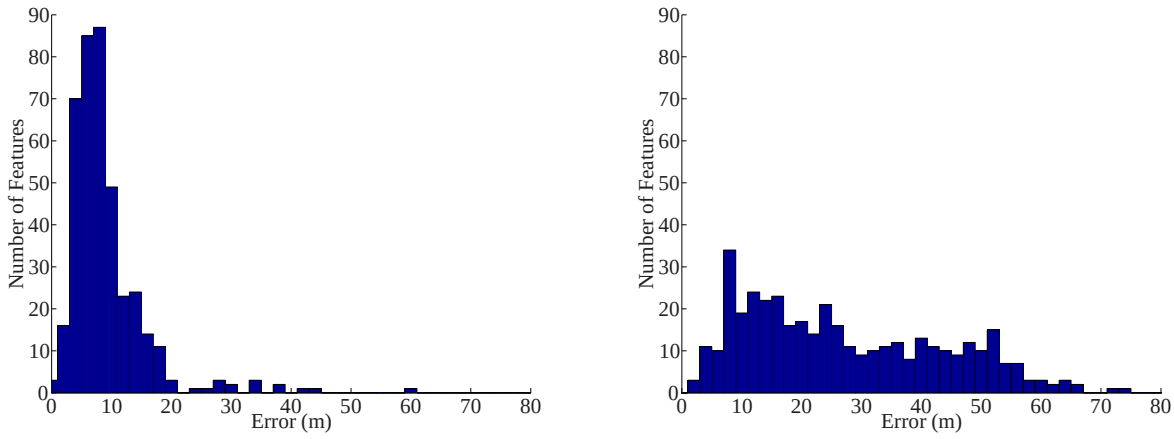



Figure 4. UAV trajectory error




(a) (b)

Figure 5. Position estimation error for each feature. Figure (a) shows the errors histogram for GPS, IMU and Camera sensor fusion and Figure (b) for IMU and Camera

van Klinken, R.D., Shepherd, D., Parr, R., Robinson, T.P. and Anderson, L., 2007. "Mapping mesquite (prosopis) distribution and density using visual aerial surveys". *Rangeland Ecology Management*, Vol. 60, pp. 408–416.
Welch, G. and Bishop, G., 2001. "An introduction to the Kalman filter". SIGGRAPH.

## 7. Responsibility notice

The authors are the only responsible for the printed material included in this paper.