Sensor Fused Three-dimensional Localization Using IMU, Camera and LiDAR

Hanieh Deilamsalehy
Department of Electrical & Computer Engineering
Michigan Technological University
Houghton, MI, USA
hdeilams@mtu.edu

Timothy C. Havens
Department of Electrical & Computer Engineering
Department of Computer Science
Michigan Technological University
Houghton, MI, USA
thavens@mtu.edu

***Abstract*—Estimating the position and orientation (pose) of a moving platform in a *three-dimensional* (3D) environment is of significant importance in many areas, such as robotics and sensing. In order to perform this task, one can employ single or multiple sensors. Multi-sensor fusion has been used to improve the accuracy of the estimation and to compensate for individual sensor deficiencies. Unlike the previous works in this area that use sensors with the ability of 3D localization to estimate the full pose of a platform (such as an unmanned aerial vehicle or drone), in this work we employ the data from a 2D *light detection and ranging* (LiDAR) sensor, which can only estimate the pose in a 2D plane. We fuse it in an *extended Kalman filter* with the data from camera and inertial sensors showing that, despite the incomplete estimation from the 2D LiDAR, the overall estimated 3D pose can be improved. We also compare this scenario with the case where the 2D LiDAR is replaced with a 3D LiDAR with similar characteristics, but the ability of complete 3D pose estimation.**

Keywords—Extended Kalman Filter (EKF); LiDAR; camera; sensor fusion; localization

I. Introduction

Pose estimation is a non-separable part of applications such as vehicle control, mapping, and *Simultaneous Localization and Mapping* (SLAM). Various sensors are commonly used to estimate a robot's pose. *Inertial measurement units* (IMU), camera, and LiDAR are among the most popular sensors for indoor localization [1]. Between these sensors, LiDAR has received less attention for 3D pose estimation. This is because 2D LiDARs can not give a complete 3D pose estimation and 3D LiDARs are bulky and expensive.

In outdoor localization, *global positioning system* (GPS) signal is usually available and since the position can be precisely obtained using GPS, it is the common choice for fusion with the data obtained from IMU or dead reckoning [2]. However, in a GPS-denied environment, such as inside a building, other sensors should be used to correct the IMU estimations. A common approach addressing this problem is IMU-camera fusion [3,4,5,6].

Unlike previous works that only use sensors which can output a full 3D pose estimation, in this paper we fuse the pose obtained from a 2D LiDAR with the data from an IMU and a camera in an extended Kalman filter (EKF) to estimate the position and orientation of an *Unmanned Aerial Vehicle* (UAV) and demonstrate that it can improve the pose estimated only by IMU and camera. Furtheremore, we replace the 2D LiDAR with a 3D liDAR with similar characteristics and compare the results with the previous scenarios.

II. Employed Sensors

A. Inertial Measurement Unit (IMU) Senors

An IMU is a device which measures the velocity, orientation, and gravitational forces using a combination of data from accelerometers, gyroscopes, and magnetometers. In this work, only the raw data from accelerometer and gyroscope are used. The 3-axis accelerometer measures the acceleration along each of the *x*, *y,* and *z* axes in the IMU local frame and the 3-axis gyroscope measures the angular velocities related to each axis with respect to the global frame.

In order to model the IMU, its bias and noise should be considered. The bias is represented by *b* and the noise by *n*. The noise is assumed to be a zero-mean white Gaussian signal and the bias is modeled as a random process. Therefore, the true and the measured values for angular velocities and linear accelerations can be written as

$$\omega_t = \omega_m - b_\omega - n_\omega, \tag{1a}$$

$$\mathrm{a_t} = \mathrm{C(q_t)(a_m - b_a - n_a) + g}, \tag{1b}$$

where the subscripts *t* and *m* respectively denote the true and the measured values, *g* is the gravity vector, q_t is the quaternion equivalent of the true attitude, and $C(q_t)$ is the rotation matrix of the true attitude q_t. In (1a) and (1b), n_ω and n_a represent the noise of gyro and accelerometer and b_ω and b_a denote the relevant biases. As mentioned, these biases are random processes and their dynamics are modeled as $\dot{b}_\omega = n_{b_\omega}$ and $\dot{b}_a = n_{b_a}$. As a result of these biases, IMU measurements tend to drift with time. To avoid this drift, IMUs are typically fused with other sensors, such as camera or LiDAR.

B. Camera Sensor

The most widely used model is the pinhole model, which describes the relationship between the coordinates of a 3D point and its projection onto the image plane of the camera. The camera model is represented by the following [7],

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{f}{Z} \begin{bmatrix} k_u & 0 & \frac{u_0}{f} \\ 0 & k_v & \frac{v_0}{f} \\ 0 & 0 & \frac{1}{f} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \tag{2}$$

where u and v are the point of interest coordinates in the image plane, X, Y and Z are the same point coordinates in 3D world frame, and k_u and k_v are the scale factors along the axes of pixel coordinates [8]. u_0 and v_0 are the pixel coordinates of the principal point and f is the focal length of the camera lens.

C. LiDAR Sensor

LiDAR is an optical remote-sensing technology that measures the distance and angle from the sensor to an object. The distance between the sensor and the object is calculated by measuring the time interval between an emitted laser pulse and reception of the reflected pulse. This technology is used to create high resolution maps for different applications. One substantial application is SLAM, in which the goal is to construct a map of an unknown environment while simultaneously keeping track of the robot's pose.

A 2D LiDAR radially scans the environment in a plane. In order to obtain the coordinates of each scanned point, the equations $x = d\cos(\phi)$ and $y = d\sin(\phi)$ are employed in which d is the distance from LiDAR to the scanned point and ϕ is the beam angle. It should be noted that x and y are the coordinates of the scanned point in the local frame of the LiDAR and not with respect to the global reference frame. In order to find the global coordinates, the local coordinates have to be transformed to the global frame.

The operation of a 3D LiDAR is similar to that of the 2D one except that instead of one measurement plane, it has two perpendicular rotational motors allowing measurement in multiple planes. As a result, a complete 3D pose can be estimated from its data.

III. Sensor Fusion Using EKF

Extended Kalman Filter (EKF) is a well known method that linearizes the state of a system about an estimate of the current mean and covariance. This can be done only if the linearizion errors are small in the update time interval. In our work, the assumption is that the sensors' frequencies are high enough and, hence, the time interval is short enough that a linear estimation is a valid hypothesis. Furthermore, EKF assumes additive process and measurement noises. The general formulation of EKF is

$$x_k = f(x_{k-1}, u_{k-1}) + \omega_{k-1} \quad , \quad z_k = h(x_k) + \upsilon_k, \tag{3}$$

where x is the state vector and z is the measurement. Subscript k denotes the time step, ω and ν are the process and observation noises, and both are zero-mean Gaussian distributed noises with covariances Q and R, assumed to be constant in this paper. In (3), u is the input control vector.

The functions f and h are usually nonlinear and cannot be used directly. Hence, the Jacobean of these functions are employed in the EKF. The filter consists of two major steps: predict and update. These can be summarized as

Predict:

$$\hat{x}_{k|k-1} = f(\hat{x}_{k-1|k-1}, u_{k-1}), \quad \text{Predict the state estimate} \tag{4a}$$

$$P_{k|k-1} = F_{k-1} P_{k-1|k-1} F_{k-1}^T + Q_{k-1}, \quad \text{Predict covariance estimate} \tag{4b}$$

Update:

$$\tilde{y}_k = z_k - h(\hat{x}_{k-1|k-1}), \quad \text{Measurement residual} \tag{5a}$$

$$S_k = H_k P_{k|k-1} H_k^T + R_k, \quad \text{Innovation} \tag{5b}$$

$$K_k = P_{k|k-1} H_k^T S_k^{-1}, \quad \text{Kalman gain} \tag{5c}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \tilde{y}_k, \quad \text{Update state estimate} \tag{5d}$$

$$P_{k|k} = (I - K_k H_k) P_{k|k-1}, \quad \text{Update covariance estimate} \tag{5e}$$

where $F_{k-1} = \frac{\partial f}{\partial x}|_{\hat{x}_{k-1|k-1}, u_{k-1}}$ and $H_k = \frac{\partial h}{\partial x}|_{\hat{x}_{k|k-1}}$.

In our work, the UAV's complete pose is defined with a translation *(x y z)* and a rotation *(α β γ)* where *α*, *β,* and *γ* represent *yaw*, *pitch* and *roll* and are defined as follows. The *yaw* angle is defined as a rotation about the z axis and *pitch* and *roll* are respectively rotations about the y and x axes.

The filter state is a 16×1 vector represented by $X = [p\ v\ q\ b_\omega\ b_a]^T$. It contains the position of the IMU (with respect to the global frame) p, the IMU linear velocity v, and the attitude q, which represents the rotation of the IMU from the reference frame. The attitude is described in the form of the quaternion which leads to a four elements vector. b_ω and b_a are gyroscope and accelerometer biases along each axis. The hypothesis is that at each time step, the change in the attitude is small; thus we can use the small angle assumption, estimating the four quaternion values by three angles. Enthusiastic readers can refer to [9] for more details.

The differential equations that govern the system state are

$$\dot{p}_t = v_t, \tag{6a}$$

$$\dot{v}_t = C(q_t)(a_m - b_a - n_a) + g, \tag{6b}$$

$$\dot{q}_t = \frac{1}{2} q_t \otimes (\bar{\omega}_m - \bar{b}_\omega - \bar{n}_\omega), \tag{6c}$$

$$\dot{b}_{\omega_t} = n_{b_\omega}, \quad \dot{b}_{a_t} = n_{b_a}, \tag{6d}$$

where subscript t represents the real values and $\bar{\omega}_m = \begin{bmatrix} 0 & \omega_m \end{bmatrix}$, $\bar{b}_\omega = \begin{bmatrix} 0 & b_\omega \end{bmatrix}$, and $\bar{n}_\omega = \begin{bmatrix} 0 & n_\omega \end{bmatrix}$.

IV. Simulation

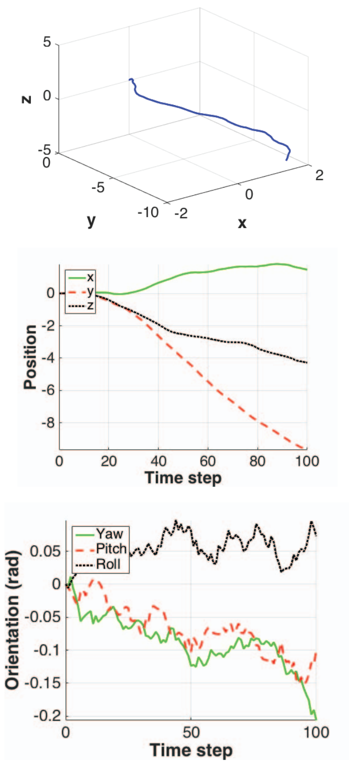
Realistic UAV flight paths were simulated in MATLAB to evaluate the performance of our algorithm. These paths were generated randomly using (1a) and (1b) and one of them is shown in Fig. 1 along with the true position and orientation of the UAV for this path. Three different scenarios were considered and compared; one is a method that previous works have employed and the two others are novel techniques proposed in this paper.

algorithm. Finally the last simulated scenario is the one that the 2D LiDAR has been replaced with a 3D one. The estimation errors are calculated and compared with respect to the true position and orientation. The simulation error results are summerized in Table 1.

TABLE I. SIMULATED PATH ERRORS FOR THE THREE SCENARIOS

Pose		Error		
		IMU & Camera	***IMU & Camera & 2D LiDAR***	***IMU & Camera & 3D LiDAR***
Position (m)	x	0.1572	0.0298	0.0218
	y	0.1260	0.0297	0.0101
	z	0.1999	0.0900	0.0246
Orientation (rad)	yaw	0.0022	0.0013	0.0012
	pitch	0.0031	0.0026	0.0016
	roll	0.0037	0.0024	0.0012

It is shown that our proposed method that employs the LiDAR has smaller error than the previous works that only fuse camera and IMU. The camera is slower than the LiDAR sensor and therefore, LiDAR can update the filter state at a higher frequency rate. This is the first reason that employing the LiDAR unit can help to improve the pose estimation. Moreover, the frequency of the LiDAR is different than the camera, thus it updates the filter at different time stamps than the camera, increasing the overall update rate of the filter. Finally, as mentioned, the LiDAR gives a more accurate pose estimation in its local plane compared to the camera.



(a) Trajectory (m)

(b) True position (m)

(c) True orientation (rad)

Fig. 1. UAV simulated path

The first experiment demonstrates the popular technique of fusing the camera estimation with the data obtained from IMU. In the second experiment, we will fuse the estimation from a 2D LiDAR with the estimation from the IMU and camera and show that, despite the fact that a 2D LiDAR can only estimate the pose in its scanning plane, the fusion with the 2D LiDAR still can improve the overall complete 3D pose estimation. Finally, for the last experiment, we will replace the 2D LiDAR with a 3D LiDAR with similar characteristics and compare the results with the previous circumstances to see how much a complete estimation from the LiDAR can improve the results.

The three sensors (IMU, camera and LiDAR) form a rigid body; therefore, the transformation between them is constant. All sensor noises are assumed to be Gaussian. To mimic realistic sensor measurements, the LiDAR estimation is more precise than the estimation obtained from the camera. The 2D LiDAR setup is assumed to be horizontal (in the local frame), which means it is able to measure the position in its local xy plane. Therefore, no estimation of z can be obtained in the LiDAR local coordinate system. Likewise, it is able to measure the yaw angle, but not the pitch or roll.

V. RESULTS AND CONCLUSION

Position estimates from the camera and LiDAR are used in the EKF as a measurement to correct the IMU pose estimation. In order to evaluate our proposed method, the mean errors of the three different sensor combinations are compared. The first one is the popular 3D pose estimation from IMU and camera fusion. In the second scenario, in addition to the IMU and camera, a 2D LiDAR is also included in the estimation

REFERENCES

[1] Del Rosario M, Lovell N, Redmond S. Quaternion-based Complementary Filter for Attitude Determination of a Smartphone. IEEE Sensors Journal, Jun 2016.

[2] Caron F, Duflos E, Pomorski D, Vanheeghe P. GPS/IMU data fusion using multisensor Kalman filtering: introduction of contextual aspects. Information Fusion. 2006 Jun 30;7(2):221-30.

[3] Mirzaei FM, Roumeliotis S. A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation. Robotics, IEEE Transactions on. 2008 Oct;24(5):1143-56.

[4] Hesch J, Mirzaei FM, Mariottini GL, Roumeliotis S. A 3d pose estimator for the visually impaired. InIntelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on 2009 Oct 10 (pp. 2716-2723). IEEE.

[5] Chambers A, Scherer S, Yoder L, Jain S, Nuske S, Singh S. Robust multi-sensor fusion for micro aerial vehicle navigation in GPS-degraded/denied environments. InAmerican Control Conference (ACC), 2014 2014 Jun 4 (pp. 1892-1899). IEEE.

[6] Hesch JA, Kottas DG, Bowman SL, Roumeliotis SI. Camera-IMU-based localization: Observability analysis and consistency improvement. The International Journal of Robotics Research. 2013 Nov 13:0278364913509675.

[7] Nemra A, Aouf N. Robust airborne 3D visual simultaneous localization and mapping with observability and consistency analysis. Journal of Intelligent and Robotic Systems. 2009 Aug 1;55(4-5):345-76. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[8] Nützi G, Weiss S, Scaramuzza D, Siegwart R. Fusion of IMU and vision for absolute scale estimation in monocular SLAM. Journal of intelligent & robotic systems. 2011 Jan 1;61(1-4):287-99.

[9] Deilamsalehy H, Havens T, Manela J, Heterogeneous Multi-Sensor Fusion for Mobile Platform 3D Pose Estimation, Journal of Dynamic Systems Measurement and Control, ASME, 2016.