Combined RGBD-Inertial based State Estimation for MAV in GPS-denied Indoor Environments

Dachuan Li, Qing Li, Nong Cheng, Qinfan Wu, Jingyan Song
Department of Automation
Tsinghua University
Beijing, China
ldc08@mails.tsinghu.edu.cn

Liangwen Tang
Science and Technology on Aircraft Control Laboratory
Flight Automatic Control Research Institute
Xi'an, China
handpl@163.com

***Abstract*— This paper presents a integrated navigation approach for state estimation of a micro aerial vehicle (MAV) that is capable of autonomous flight in GPS-denied, indoor environments. The solution combines RGB-D sensor and inertial sensors in a tight-coupling navigation manner. Motion estimates from RGB-D visual odometry and inertial measurements are fused using an improved Extended Kalman Filter-based fusion algorithm to provide an accurate estimate of the relative position, velocity and attitude. Instead of using a global reference frame, a view-based map is employed and the algorithm maintains the position and heading relative to the current map node in the fusion algorithm. In addition, a closed-form covariance is developed to qualify the uncertainty of the RGBD visual odometry measurements, which is utilized for state update of the navigation filter. Our approach allows efficient measurement updates and enables the incorporation of RGBD visual odometry uncertainty. Experimental results of a quadrotor MAV flying in a GPS-denied indoor environment demonstrate the performance of the proposed approach. Comparisons of state estimates with ground truth measurements are also provided.**

Keywords—State estimation; RGBD-IMU data fusion; visual odometry; RGB-D sensor; micro aerial vehicle

I. INTRODUCTION

Micro Aerial Vehicles (MAVs) are capable of accomplishing many military and civilian tasks without taking the risk of human life. In recent years, there has been an increasing interest on developing MAVs that can autonomously operate in indoor or dense urban areas, thereby enabling MAVs to accomplish a wider range of task scenarios such as indoor reconnaissance and surveillance, collapsed building exploration, earthquake and fire disaster relief, etc..

Since the stabilization, precise control and path planning of MAVs require accurate and fast estimation of angular rate, attitude, velocity and position, accurate state estimation and localization is one of the key enabling technologies for MAV indoor flight in GPS-denied environments. However, there exists significant technical challenges to ensure reliable state estimation for MAVs. Payload limitations force MAVs to rely on low-cost, lightweight MEMS(Micro Electro Mechanical Systems)-based IMUs with unsteady sensor bias and unbounded drift, thus it is infeasible to obtain position and velocity estimates by integrating forward the acceleration and angular measurements of these IMUs as the estimates tend to drift rapidly. As a result, in order to obtain more accurate estimates, it is a common solution to fuse inertial measurements with absolute position information provided by external navigation aids such as GPS[1] and camera-based motion capture systems[2] (e.g. Vicon system).

While most current navigation algorithms for MAVs operating in outdoor environments rely heavily on GPS information, it is not suitable for indoor flight since GPS signals are normally unreliable or even unavailable in indoor or dense urban environments. Moreover, actual application scenarios are also without access to external camera systems. As a result, the combined navigation of MAVs must only rely on on-borad IMUs and exteroceptive sensors such as laser range finders and cameras .

Combining visual and inertial sensors with inertial has proved to be viable and effective for MAV indoor navigation, and there has been considerable research on developing combined visual-inertial state estimation algorithms for MAVs. Previous work in [3] ,[4] and [5] leveraged 2D laser range finders for small quadrotors performing indoor navigation and SLAM. These systems utilize laser-scan matching algorithms and combine the data with IMU measurements in a EKF filter to provide sate estimation. Girish C. et al. present their self-contained quadrotor navigation system in [6], which uses scanning laser rangefinder and a extension of EKF based SLAM algorithm to provide position and heading estimate, while [7] implements a Gussian Particle Filter-based approach to combine inertial information and laser rangefinder measurements for a fixed-wing vehicle flying in GPS-denied indoor environments.

In addition to laser and inertial–based navigation for MAV indoor flight, there has been a variety of research on visual aided state estimation using either monocular or stereo vision. [8] demonstrates the utility of stereo vision on a quadrotor for indoor flight, an EKF filter fuses the stereo visiual and inertial measurements to obtain an estimate of vehicle's position, velocity and acceleration. Similarly, M. Acgtelik et al. have also developed a state estimation method that combines stereo visual odometry, laser odometry and IMU measurements [4]. [9] presents a robust ego-motion estimation method in challenging industrial environments, which uses inertial information to aid the feature matching of stereo vision. In

This work is supported through Aviation Science Foundation of China. （ 20100758002）

978-1-4673-5769-2/13/$31.00 ©2013 IEEE

addition, several researchers have also attempted to develop approaches using monocular vision, examples of such combined monocular vision-inertial estimation approaches can be found in [10] and [11].

Although exteroceptive sensors such as laser range finders and cameras have been successfully implemented for indoor motion estimation of MAVs, laser and visual based odometry usually require specific assumptions about the structures and features of environments that limit their effectiveness only to certain scenarios. Since laser rangefinders only provide distance measurements along a plane around the sensor, the scan-matching based approach is only useful in environments characterized by unique and vertical structures, and it may fail around certain homogeneous structures such as corridors. In addition, laser rangefinders are unable to perceive height and range information outside the 2D sensing plane, which makes them less effective in complex 3D scenes. In contrast, while camera sensors can make use of richer 3D information of the environment, vision based methods require the environments contain distinctive features and the processing algorithms are usually computationally demanding. Moreover, since visual odometry usually has unbounded global drift, it is generally integrated with SLAM algorithms to bound the estimate error. However, the loop closure process is generally computationally intensive, therefore it is not feasible for onboard implementation due to limited computation capabilities.

To tackle these problems, this paper proposes a relative state estimation approach for MAVs operating in GPS-denied indoor environments by combining an onboard RGBD sensor and IMUs. The RGBD sensor adopted in our system is the Microsoft Kinect(see Fig.1), which can provide RGB color image with depth data per pixel. A RGBD-based robust motion estimation algorithm is developed for estimating the vehicle's states by calculating the relative translation and rotation, the state estimates are then fused with IMU measurements using an extended EKF-based filter to provide fast and reliable estimation of MAV's position and velocity.

In contrast to previous implementations which utilize global states for estimation, our state estimation approach employ a view based graph consisting of nodes(keyframes with sensor measurements and associated position and heading) and edges(estimated transformation between nodes). During real time operations, navigation states used by the filter are kept relative to the current reference keyframe in the graph, which is replaced by a new declared keyframe once the overlap between the current view and reference keyframe diminishes. This approach avoids additional transformation of relative states in the filter. In addition, unlike conventional SLAM frameworks that directly solve global optimization using time-consuming loop closing techniques, the relative navigation architecture incorporates a sparse pose adjustment approach to eliminate global drifts and improves global consistency of the estimates from the RGBD odometry. This optimization process runs in a background manner, thereby making the algorithm more flexible and efficient.

In addition, a novel closed-form covariance is developed in this paper for determining the uncertainty of estimates from RGBD odometry. The RGBD measurement uncertainty is further incorporated into the state estimation filter to update the full state estimate. Therefore, this covariance provides a proper way for statistically qualifying the confidence of RGBD measurements and analyzing the influence of environment as well as sensor noises on RGBD based state estimates. The proposed state estimation approach can provide accurate and reliable state estimates, enabling the MAV to operate in GPS-denied indoor environments without relying on any external navigation aids. Finally, we have implemented our algorithm on a quadrotor and evaluate its effectiveness through indoor flight tests.

This paper is organized as follows: Section Ⅱ gives an overview of the navigation system and the state estimation framework, and a detailed description of RGBD-based motion estimation method is presented in Section Ⅲ. The combined state estimation is presented in Section Ⅳ. After the experimental results in Ⅴ, the paper is concluded in Section Ⅵ.

II. System Framework Overview

The architecture of the MAV navigation system is illustrated in Fig.2. As can be seen from the diagram, the navigation system consists of three primary sensors: a Microsoft Kinect, an IMU module and a sonar altimeter. The Microsoft Kinect is capable of providing a 640×480 RGB color image with depth measurements at 30Hz. The IMU module consists of a three-axis gyroscope, a three-axis accelerometer and a three-axis magnetometer. An ultrasonic range finder is equipped for altitude measurement. Measurements from the IMU and Kinect based odometry are fused using the data fusion filter (implemented on a onboard embedded computer) to form an accurate estimation of the MAV's 6-DOF states, which will be described in the following sections.

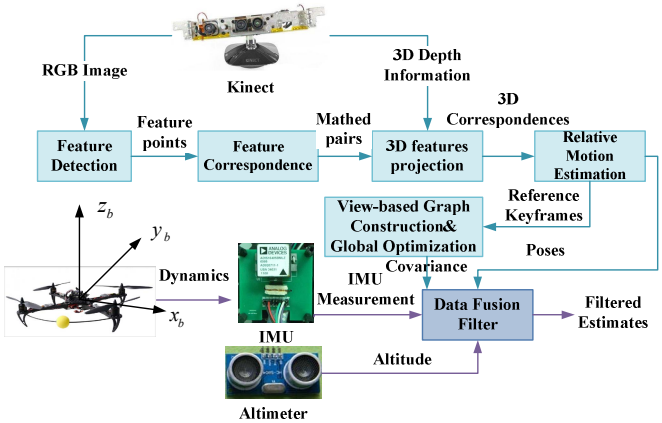

Fig. 1 Schematic of combined RGBD-Inertial navigation system

III. RGB-D based Real Time Motion Estimation

Estimating the MAV's motion from RGB-D data consists of incrementally determining the relative ego-motions at each time step by aligning consecutive RGB-D frames, a procedure which is also referred to as visual odometry. The process of our robust RGB-D odometry is outlined in Fig.1. The RGB-D odometry approach consists of a chain of subroutines. First, features of interest are identified and extracted from the RGB-

D image, these features are then matched across frames, enabling us to project these 2D matched pairs into 3D space using depth data to find 3D correspondence. Finally, a robust motion estimation method is used to compute the MAV's motion in 6 DOF. In addition, a global estimate optimization is developed to correct global drift and construct the relative navigation graph. Each of these sub-processes will be described in detail in the following parts.

A. Feature Detection

The raw RGB-D image acquired by the Kinect is first converted to grayscale and preprocessed for feature detection. Although there is a number of options available for feature extraction (including Harris corner[12], SIFT[13], FAST[14], etc.), we adopted the SURF(Speeded Up Robust Features)[15] detector considering its robustness to variations of scale, rotation, as well as affine distortion noise and image blur which can be caused by the motions of MAV. At each time step, the feature detection process yields a sufficient number of features that are distinctly recognizable from the images, such that finding the correspondence between features of different images is viable. Examples of features extracted from Kinect images are shown in Fig.2.

B. Frame to Frame Feature Correspodence

After the feature detection procedure, each detected feature is assigned a descriptor consisting of 128 floating points which represent the brightness values of pixels around the feature. Features in different images are then matched by comparing their descriptor values, using a k-nn search strategy. We use Euclidean distance as the metric of the k-nn search and $k = 2$. Define d_1, d_2 as the Euclidean distances of descriptors between each feature and its best match, as well as its second best match, respectively. A feature match is declared when two features fulfill the following constraints:

$$d_1 < \varepsilon_1, d_1 / d_2 < \varepsilon_2 \tag{1}$$

where $\varepsilon_1, \varepsilon_2$ denote thresholds and we set $\varepsilon_1 = 0.5, \varepsilon_2 = 0.7$ in our algorithm.

Fig. 2Detected features and mathes between two consecutive images

In order to reduce the rate of incorrect feature matches, a RANSAC procedure [16] is employed to select final matching results and prune out the incorrect correspondences. Examples of the feature matching results are presented in Fig.3

C. Robust Motion Estimation

Each feature detection and matching step yield two sets of corresponding image features from RGB images of consecutive time steps, during which the MAV has undergone a rotation R and a transformation t. These image features are firstly projected into 3D space using their corresponding 3D depth data from the Kinect depth image, thereby yielding two sets of matched 3D features. With these 3D correspondences, the relative motion (R and t) between the previous and current time step can be calculated using a robust least-square method.

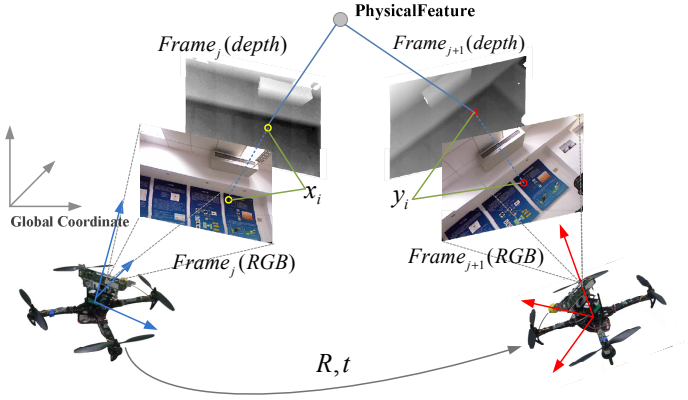

Fig. 3 Relative motion estimation procedure. The MAV's motion can be decomposed into rotation and translation. And $x_i y_i$ denote examples of a3D corresponding feature pair from consecutive frames (before and after the MAV has undergone R and t).

As is illustrated in Fig.3, the motion estimation algorithm must solve the following problem: Given two sets of 3D correspondences, denoted as $\mathbf{X} \in R^3 (\mathbf{X} = \{\mathbf{x}_i\}, i = 1...n)$ and $\mathbf{Y} \in R^3 (\mathbf{Y} = \{\mathbf{y}_i\}, i = 1...n)$, find the optimal transformation $\Delta \in SO(3), \Delta = \{R, t\}$ such that by applying Δ to $\mathbf{X}$ ($\Delta \otimes \mathbf{X} = R\mathbf{X} + t$),the squared error $\varepsilon(R,t)$ is minimized, $\varepsilon(R,t)$ is given by:

$$E(R,t) = \frac{1}{n} \sum_{i=1}^{n} \left\| \mathbf{y}_i - (R\mathbf{x}_i + t) \right\|_2^2 \tag{2}$$

where n denotes the number of 3D correspondences. In order to achieve better robustness to noises, a RNASAC based procedure is employed to calculate the relative motion. The overall robust motion estimation algorithm is presented in Algorithm.1.

Algorithm1: Robust Motion Estimation

Input: $X = \{x_i\}, Y = \{y_i\}, i = 1...n$
Output: Rotation R and translation t that minimize $\varepsilon(R,t)$
Begin:

$k \leftarrow 0$
while $k<max$ **do**
  select subset $S^{(k)} = \{(\mathbf{x}_1^{(k)}, \mathbf{y}_1^{(k)}), (\mathbf{x}_2^{(k)}, \mathbf{y}_2^{(k)}), (\mathbf{x}_3^{(k)}, \mathbf{y}_3^{(k)})\}$ from $\mathbf{X}$, $\mathbf{Y}$
  compute $R_0^{(k)}, t_0^{(k)}$ and Euler angle $\theta_0^{(k)}, \phi_0^{(k)}, \psi_0^{(k)}$ using $S^{(k)}$
  set $S_c^{(k)} \leftarrow S^{(k)}$
  for each $p_i^{(k)} = (\mathbf{x}_i^{(k)}, \mathbf{y}_i^{(k)}) \in \{(\mathbf{x}_4^{(k)}, \mathbf{y}_4^{(k)}), \cdots, (\mathbf{x}_n^{(k)}, \mathbf{y}_n^{(k)})\}$ **do**
    add $p_i^{(k)}$ to $S^{(k)}$, $S_t^{(k)} \leftarrow S^{(k)} \cup \{p_i^{(k)}\}$
    compute $R_i^{(k)}, t_i^{(k)}$ using $S_t^{(k)}$
    extract Euler angle $\theta_i^{(k)}, \phi_i^{(k)}, \psi_i^{(k)}$ from $R_i^{(k)}$, compute $\left\| t_i^{(k)} - t_0^{(k)} \right\|_1$
    if $\left|\theta_i^{(k)} - \theta_0^{(k)}\right| < \varepsilon_1$, $\left|\phi_i^{(k)} - \phi_0^{(k)}\right| < \varepsilon_2$, $\left|\psi_i^{(k)} - \psi_0^{(k)}\right| < \varepsilon_3$ and $\left\| t_i^{(k)} - t_0^{(k)} \right\|_1 < \varepsilon_4$
    then add $p_i^{(k)}$ to $S_c^{(k)}$
    end

```
    remove p_i^(k) from S^(k)
  end
  k ← k+1
end
k ← arg min sizeof(S_c^(k))
compute the final result of R, t using S_c^(k)
output R, t
```

In line 4, 8, 18 of each time step, the optimal R and t is calculated using the singular value decomposition(SVD) method [8]. Algorithm 1 generates the rotation and translation between two sets of 3D corresponding points with respect to a fixed coordination. However, in this approach the sets of points are fixed and the camera coordinate (e.g. the MAV body coordinate frame) is in motion. Therefore, the relative rotation ΔR and translation Δt of current frame to MAV's previous body frame is given by:

$$\Delta R = R^T,\ \Delta t = -R^T t \tag{3}$$

The MAV's pose at a given step relative to an initial pose T_0 can be obtained by performing a chain of homogeneous transformations:

$$T_c = T_0 \cdot \Delta T_{t-n+1} \cdot \ldots \cdot \Delta T_{t-1} \cdot \Delta T_t,\ T_p = \Delta T_{t-n+1} \cdot \ldots \cdot \Delta T_{t-1} \tag{4}$$

where

$$T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \tag{5}$$

denotes a 4×4 transformation matrix. Therefore, the transformation update over time steps is performed by right multiplication of a single latest transformation matrix. And the Euler angles(θ, ϕ, ψ) of the MAV can be calculated from translation R as follows:

$$\begin{aligned} \theta &= \mathbf{atan2}(-r_{31}, \sqrt{r_{11}^2 + r_{12}^2}) \\ \phi &= \mathbf{atan2}\left(\frac{r_{32}}{\cos(\theta)}, \frac{r_{33}}{\cos(\theta)}\right) \\ \psi &= \mathbf{atan2}\left(\frac{r_{21}}{\cos(\theta)}, \frac{r_{11}}{\cos(\theta)}\right) \end{aligned} \tag{6}$$

where θ, ϕ, ψ denote the pitch, roll and yaw attitude angle of MAV, respectively, and r_{ij} represents specific element of R.

D. Drift correction and global estimate optimization

Since small measurements errors in each estimate step will accumulate over time, resulting in significant and unbounded global position drift over time. An optimization approach is additionally employed to reduce short-term error and long-term global drift, which consists of two sub-processes.

In order to eliminate short scale error, the relative motion is computed by comparing 3D correspondences in successive frames I_i (i=k+1...k+n)against those in the reference keyframe I_r^k. Once there are not sufficient correspondences between the current image I_c and I_r^k, I_c will be declared as the new keyframe I_r^{k+1}, as depicted in Fig. 4. This can reduce the motion estimation drift between successive frames.

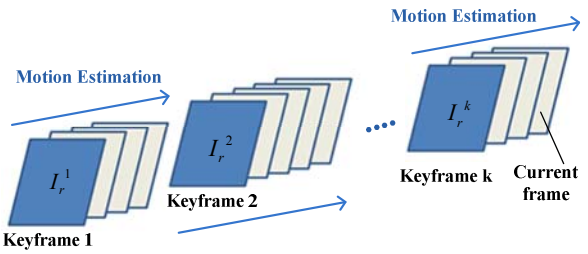

Fig. 4 Keyframe method for drift elimination. Dark blue colored blocks denote keytframes declared in different time steps

In addition, a global optimization approach is employed to bound long term drift and improve global consistency by constructing a relative pose graph consisting of pose nodes that represent MAV's poses and associated frames (Details of the pose graph concept will be described in section Ⅳ, A), and nodes are connected by edges denoting motion constraints. Once loop closure is detected, additional constraints are imposed between disjoint nodes, and finally the optimized poses are obtained by solving the optimization problem using a Levenberg-Marquardt approach[16].

IV. Combined State Estimation Algorithm

A. View-Graph based Relative Navigation

A node-edge based view graph is constructed as the MAV compares its current frame against the reference key frame and estimates the relative motion, as described in previous sections. Every node in the graph represents a key frame (with image and depth measurements taken at that pose) and the associated pose (3D position and yaw angle), and the edge connecting two adjacent nodes denotes the estimated transformations between poses. The current states of the MAV are kept relative to the local frame of the current reference node that associated to the reference key frame. Moreover, global states of the MAV relative to the global coordinate can be obtained by summing up the vectors of the edges and the current states. The relative navigation graph is showed in Fig.6.

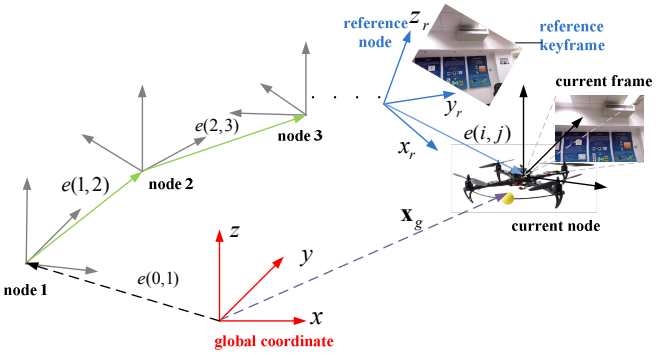

Fig. 5 Relative navigation graph

The graph-based navigation framework has the following features:

1)It allows the straightforward integration of the relative pose estimates from RGBD odometry in the EKF-based fusion framework. In contrast to practices that uses global referenced states, the states in the fusion filter are instead kept relative to the current reference node that associated to the reference key

frame. Therefore, the visual measurements can be directly incorporated in the EKF filter for measurement update without additional transformation from relative states to global updates.

2)Felxibility in transformation from relative estimates to global states. Since states are kept with respect to the current reference node frame, the global states can be derived, if desired, by traversing the edges in the graph. As depicted in Fig.5, the current state with respect to the global frame (denoted as $\mathbf{x}_g$) can be obtained by summing the edges (denoted as $e(i, j)$) and the current state.

3)It supports the integration of back-ends global optimization without impacting the real time performance of the relative motion estimation and data fusion . Since RGBD odometry has unbounded global drift caused by accumulation of single step errors, a global optimization using sparse pose adjustment method is performed as background process that provides corrections for the global errors of the estimates and improves global consistency, when loop closure is detected. This optimization runs in an opportunistic manner such that it does not interfere with the relative state estimation and data fusion process.

4)The view-graph preserves rich information for 3D map construction and path planning. Each node of the graph preserves the sensor measurements of the environment with the states in which the measurement was taken. Therefore, the information in the view-graph allows for high level tasks such as map construction and path planning。 (However, this paper focus on the state estimation using such information)

B. Analysis of the Error Covariance in RGBD Motion Estimation

It is necessary to qualify the error covariance of the motion estimation provided by the RGBD odometry, since the RGBD motion estimates will be integrated in the fusion algorithm. To do so, we qualify the covariance matrix using the Hessian based method, which is closely related to the one described in [17].

The motion estimation algorithm can be considered as a function that calculates the estimates of motion $\hat{T}$ minimizing the cost function J (squared error E defined in Eq.(2)), using the measurements $\hat{z}$ (3D points): $\hat{T} = \varphi(\hat{z}) = \arg\min_T J(\hat{z}, T)$.

which is not in closed-form. Therefore, φ can be approximated as :

$$\varphi(z) = \varphi(E[z]) + \left.\frac{\partial \varphi}{\partial z}\right|_{z=E[z]} (z - E[z]) \tag{7}$$

which can also be written as:

$$\varphi(z) - \varphi(E[z]) = \left.\frac{\partial \varphi}{\partial z}\right|_{z=E[z]} (z - E[z]) \tag{8}$$

Therefore, the covariance of T can be given as:

$$R_T = E[(\varphi(z) - E[\varphi(z)])(\varphi(z) - E[\varphi(z)])^T] = \left.\frac{\partial \varphi}{\partial z}\right|_{z=E[z]} R_z \left(\left.\frac{\partial \varphi}{\partial z}\right|_{z=E[z]}\right)^T \tag{9}$$

where R_z is the covariance matrix of z .

The estimates from the RGBD odometry can be seen as a noisy measurement of the MAV's real motion. As aforementioned, since motion estimates are obtained by matching two successive clouds of 3D points provided by the RGBD sensor, the RGB sensor noise (with the covariance R_z) is the primary source of the total error. Therefore, we derive the relationship between the covariance (denoted as R_T) of motion estimates and R_z , using the implicit function theorem.

Implicit function theorem [18]: Let $f = f(x, z)$ be a function from an open set $E \subset R^{n+m}$ to R^n , and there exists a point $(x_0, z_0) \in E$ such that $f(x_0, z_0) = 0$. Then there exists open sets $M \subset R^n$ and $N \subset R^m$ (where M , N are neighborhoods of x_0, z_0 ($z_0 \subset M$, $x_0 \subset N$),respectively), and a unique function $g : M \to N$ such that $x_0 = g(z_0)$, $f(g(z), z) = 0$ for each $z \in M$, moreover

$$\frac{\partial g(z)}{\partial z} = -\left(\frac{\partial f(g(z), z)}{\partial x}\right)^{-1} \frac{\partial f(g(z), z)}{\partial z} \tag{10}$$

In our problem, since $\hat{T}$ is the result of $\varphi(z)$ that minimizes the cost function J , we have $\partial E(\hat{z}, \hat{T}) / \partial T = 0$. Let f≒ $\partial J / \partial T$, $x_0 = \hat{T}$ and $z_0 = E[z]$, then apply the implicit theorem, we can obtain:

$$\left.\frac{\partial \varphi}{\partial z}\right|_{z=E[z]} = -\left(\frac{\partial^2 J}{\partial T^2}\right)^{-1} \frac{\partial^2 J}{\partial T \partial z} = -H^{-1} \frac{\partial^2 J}{\partial T \partial z} \tag{11}$$

where H denotes the Hessian matrix of J with respect to T. Therefore, Eq.(9) can be written as:

$$R_T = H^{-1} \frac{\partial^2 J}{\partial T \partial z} R_z \left(\frac{\partial^2 J}{\partial T \partial z}\right)^T H^{-T} \tag{12}$$

From Eq(2), the cost function can be transformed as:

$$J(z, T) = E(X, Y, R, t) = \frac{1}{n} \sum_{i=1}^{n} \left\| \mathbf{y}_i - (R\mathbf{x}_i + t) \right\|^2 = \frac{1}{n} \sum_{i=1}^{n} J_i^2 \tag{13}$$

Then from (11) and (13), we can obtain:

$$H = \frac{\partial^2 J}{\partial T^2} = \frac{2}{n} \left(\sum_{i=1}^{n} \frac{\partial J_i^T}{\partial T} \frac{\partial J_i}{\partial T} + J_i \sum_{i=1}^{n} \frac{\partial^2 J_i^T}{\partial T^2} \right) \approx \frac{2}{n} \sum_{i=1}^{n} \frac{\partial J_i^T}{\partial T} \frac{\partial J_i}{\partial T} \tag{14}$$

$$\frac{\partial^2 J}{\partial T \partial z} \approx \frac{2}{n} \sum_{i=1}^{n} \frac{\partial J_i^T}{\partial T} \frac{\partial J_i}{\partial z} \tag{15}$$

Therefore, substituting (14) and (15),Eq.(12) becomes:

$$R_T = H^{-1} \left(\sum_{i,j}^{n} \frac{\partial J_i^T}{\partial T} \frac{\partial J_i}{\partial z} R_z \frac{\partial J_j^T}{\partial T} \frac{\partial J_j}{\partial z} \right) H^{-T} \tag{16}$$

Assuming that 3D feature measurements from RGBD are corrupted to Gaussian noise with identical variance σ^2 , and are uncorrelated with each other, then the covariance matrix of z becomes $R_z = diag[\sigma^2 \cdots \sigma^2]_{n \times n}$,and (16) becomes:

$$R_T = \sigma^2 H^{-1}\left(\sum_{i,j}^{n}\frac{\partial J_i^T}{\partial T}\frac{\partial J_i}{\partial z}\frac{\partial J_j^T}{\partial T}\frac{\partial J_j}{\partial z}\right)H^{-T} \tag{17}$$

which gives the close-form representation for the uncertainty of the RGBD motion estimates. This covariance can be easily computed and is then used for the measurement update of data fusion algorithm.(The standard deviation of the noise of Kinect sensor is approximately $\sigma \approx 0.2m$)

C. EKF-based Data Fusion and State Estimation

An extension of a mixed continuous-discrete EKF architecture is used to combine RGBD motion estimates with inertial measurements, in order to correct for the drift of IMU sensors and obtain the accurate estimates of the MAV's position, velocity and attitude..

The state vector estimated by the data fusion filter is given as:

$$\mathbf{x} = \begin{bmatrix}\mathbf{s_r} & \mathbf{q} & \delta\theta & \mathbf{v} & \boldsymbol{\varepsilon} & \nabla\end{bmatrix}^T \tag{18}$$

The state vector consists of six quantities: $\mathbf{s_r} = [x_r, y_r, z_r]^T$ represents the relative position, which is with respect to the previous key frame in our relative navigation graph. $\mathbf{q} = [q_1, q_2, q_3, q_4]^T$ denotes the attitudes in unit quaternion. $\delta\theta$ is the three-dimensional error state of the quaternion, which is used to correct the attitude estimate in each measurement uppdate. $\mathbf{v} = [v_1, v_2, v_3]^T$ denotes velocity. The filter is also designed to estimate biases $\boldsymbol{\varepsilon}, \nabla$ of the gyroscope and accelerometer , respectively.

1) Propagation

The following process model is used for estimating the states :

$$\begin{aligned}
\dot{\hat{\mathbf{s}}}_r &= C_b^r \hat{\mathbf{v}} \\
\dot{\hat{\mathbf{q}}} &= \frac{1}{2}\hat{\mathbf{q}} \otimes \hat{\omega} \\
\dot{\delta\theta} &= 0 \\
\dot{\hat{\mathbf{v}}} &= C_q \hat{\mathbf{a}} + g \\
\dot{\hat{\boldsymbol{\varepsilon}}} &= 0 \\
\dot{\hat{\nabla}} &= 0
\end{aligned} \tag{19}$$

where $\hat{\bullet}$ denote estimated values. C_b^r represents the rotation matrix from body frame to the key node frame. C_q denotes the rotation matrix corresponding to q and g is the gravity vector. The attitude quaternion estimate is updated using the following equation:

$$\dot{\hat{\mathbf{q}}} = \frac{1}{2}\hat{\mathbf{q}} \otimes \hat{\omega} = \frac{1}{2}\hat{\mathbf{q}}\begin{bmatrix} 0 & -p & -q & -r \\ p & 0 & r & -q \\ q & -r & 0 & p \\ r & q & -p & 0 \end{bmatrix} \tag{20}$$

where $\hat{\omega} = [p \quad q \quad r]$ denotes the angular rates computed from the bias-corrected gyroscope measurements: $\hat{\omega} = \omega_{raw} - \hat{\boldsymbol{\varepsilon}}$. The error quaternion between the true attitude $\mathbf{q}$ and the estimated $\hat{\mathbf{q}}$ quaternion can be approximated as $\delta\mathbf{q} \approx [1 \quad \delta\theta]^T$ such that :

$$\mathbf{q} = \hat{\mathbf{q}} \otimes \delta\mathbf{q} \tag{21}$$

Whereas the acceleration $\hat{\mathbf{a}}$ is defined by the accelerometer measurements minus the bias, i.e. $\hat{\mathbf{a}} = \mathbf{a}_{raw} - \nabla$.

Using the above continuous process model, the states can be propagated forward over time. The error covariance is propagated as follows:

$$\dot{\mathbf{P}} = \mathbf{AP} + \mathbf{PA}^\mathbf{T} + \mathbf{Q} \tag{22}$$

where A denotes the Jacobian of model (19) with respect to the state vector, and Q represents the uncertainty of the process, which can be tuned in practical implementations.

2) Measurement update

The measurement update is completed using the 6-degrees of freedom relative motion estimation between the current frame and the reference key frame provided by the RGBD odometry, along with the covariance matrix $\mathbf{R}_T$ .As the relative position estimate is with respect to the coordination system of the reference key frame, the measurement model is given as:

$$\mathbf{h(x)} = \begin{bmatrix} \mathbf{s}_r \\ \mathbf{C_q q} \end{bmatrix} \tag{23}$$

where position $\mathbf{s}_r$ and attitude $\mathbf{q}$ are derived from the relative translation and rotation computed by the RGBD odometry respectively, and $\mathbf{C_q}$ denotes the transformation matrix corresponding to $\mathbf{q}$. The measurement model is used in the measurement update as follows:

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^T(\hat{\mathbf{x}}_k^-)(\mathbf{R}_T + \mathbf{H}_k(\hat{\mathbf{x}}_k^-)\mathbf{P}_k^- \mathbf{H}_k^T(\hat{\mathbf{x}}_k^-))^{-1} \tag{24}$$

where $\mathbf{K}$ denotes the Kalman gain, and $\mathbf{H}$ represents the Jacobian of the measurement model. The superscripts + and – represent the corrected and predicted variables, respectively. After the Klaman gain is computed, the covariance matrix $\mathbf{P}$ is updated using the Joseph Stabilized form:

$$\mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k(\hat{\mathbf{x}}_k^-))\mathbf{P}_k^-(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k(\hat{\mathbf{x}}_k^-))^T + \mathbf{K}_k\mathbf{R}_T\mathbf{K}_k \tag{25}$$

Finally , we the predicted states are corrected as follows:

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{z}_k - \mathbf{h}(\hat{\mathbf{x}}_k^-)) \tag{26}$$

Note that the quaternion state is updated separately as:

$$\mathbf{q}_k^+ = \hat{\mathbf{q}}_k^- \otimes \delta\mathbf{q}_k \tag{27}$$

Since part of the states in the filter are kept relative to the current reference node, these relative portions must be augmented when a new key frame (i.e. a new node in the view-gragh) is declared. The mean values of the new node are assigned to zero, and the covariance are also augmented in the

filter. In addition, the states and covariance are saved at each time step to accommodate for the delay of the states.

V. Experimental Results

We have conducted several real-time flight experiments in order to evaluate the performance of the navigation system and the combined state estimation framework. An indoor experimental environment was set up with an external motion capture camera providing ground truth measurements of the position and velocity. The RGBD odometry and data fusion algorithm were implemented on the quadrotor MAV as illustrated in Fig.6. For all flight tests, the onboard Kinect captures RGB and depth images of the environment at a rate of 30Hz, while the IMU module provide inertial measurements at a higher frequency of 100Hz.

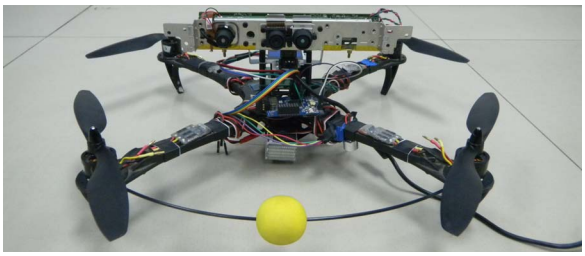

Fig. 6 MAV equipped with a Kinect RGBD and MEMS sensors

In our flight experiments, the quadrotor was controlled to follow a rectangular trajectory in the indoor environment. This scenario contains various typical indoor scenes, either poorly textured, or with sufficient features. For all experiments, the RGBD motion estimation algorithm as well as the data fusion filter ran simultaneously, and the ground truth trajectory and associated measurements were recorded from the external motion capture system. We compared state estimates of our algorithm with ground truth in order to qualitatively validate the accuracy of state estimates.

The plot of the IMU based position and velocity estimates are shown in Fig.7. These estimates are obtained solely by using measurements of the gyroscope and accelerometer. As can be found from the plot, the results demonstrates that the inertial estimates would drift rapidly without corrections, resulting in unacceptable accuracy.

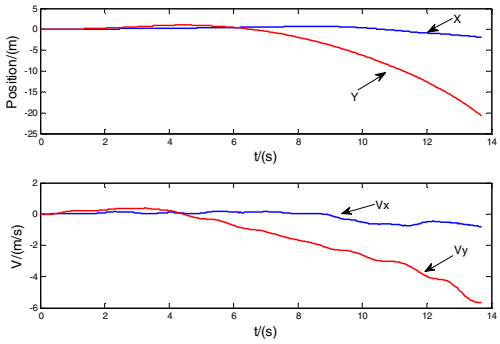

Fig. 7 Position and velocity estimates using IMU mesurements alone

The position and velocity estimates from the data fusion filter versus the ground truth velocity values (ground truth position values are shown in Fig.10.) are shown in Fig.8. As is shown in the plot, the results indicate that the proposed approach is effective in estimating the actual position and velocity of the MAV. In comparison with Fig7, these results also demonstrate that the bias of IMU measurements can be effectively corrected by combining the RGBD relative motion estimates and inertial estimates.

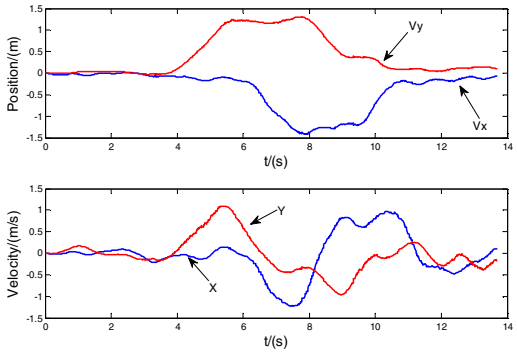

Fig. 8 Positon and velocity estimtates from data fusion

Fig 9 illustrates the plots of the Euler angle estimates (θ, ϕ, ψ). Note that the heading of the MAV (yaw angle) was kept in a limited scale such that the Kinect could get sufficient features for motion estimation during the experiments. Again, these results indicate that attitudes of the MAV can also be well estimated using our approach.

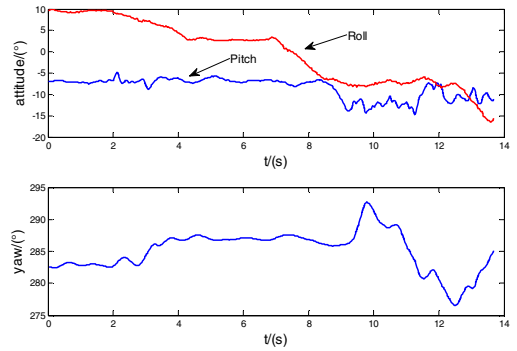

Fig. 9 Attitude estimates(pith, roll and yaw) from data fusion

An example of an estimated 3D trajectory is plotted in Fig.10, along with ground truth comparison. Compared with ground truth measurements, the position estimates closely matches the ground truth values, although the position error is slightly larger in the X (east) direction(with a maximum deviation of approximately 8cm) than that in the Y(north) direction. This is due to decrease of distinctive textures in the environment along the X direction. However, the position estimates still achieves acceptable accuracy, which is sufficient for trajectory control.

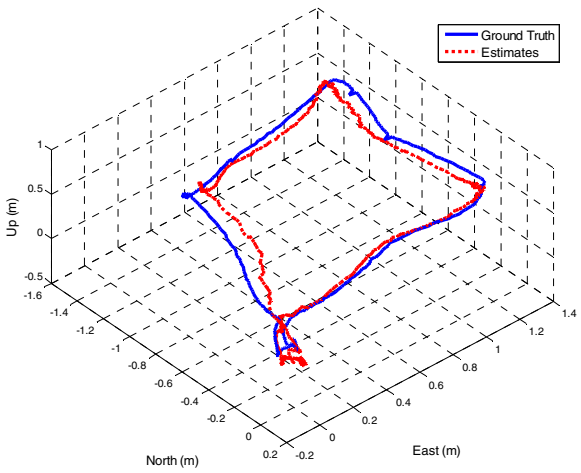

Fig. 10 Comparison between position estimates (red dotted line) of a 3D trajectory with ground truth measurements (blue solid line) from a external motion capture system.

VI. CONCLUSION AND FUTURE WORK

This paper presents a combined state estimation framework designed for MAV performing indoor exploration flights based on the inertial and RGBD sensors. A robust motion estimate approach is developed using the RGB and depth information of the environment provided by the RGBD sensor. The motion estimates from the RGBD odometry are fused with inertial measurements through an extended EKF filter to produce an accurate estimate of the MAV's 6-DOF motion states. The important feature of the proposed framework is that it incorporates a relative navigation architecture based on a view-graph, with a global optimization running opportunistically to correct for global drift of the RGBD motion estimates. In contrast to methods that utilize global states, our relative navigation architecture avoids additional transformations of the relative states and allows for more flexibility. In addition, a novel closed-form covariance of the RGBD odometry is proposed to analyze the uncertainty of the RGBD motion estimates in order to integrate these information in the data fusion algorithm. The effectiveness of our approach is validated via real time experiments on a quadrotor platform. Flight test results demonstrate that the proposed framework and associated algorithms can provide the MAV with reliable and fast state estimates, enabling the MAV to explore in GPS-denied environments without relying on external navigation aids

Our future work will focus on integrating the state estimation algorithm with control and path planning components to form a closed-loop framework for indoor flight. We are also planning to evaluate the performance of the state estimation approach in more challenging actual indoor environments.

REFERENCES

[1] Christophersen, H. B., Pickell, W. R., Neidoefer, J. C., Koller, A. A., Kannan, S. K., and Johnson, E. N., "A Compact Guidance, Navigation, and Control System for Unmanned Aerial Vehicles," Journal of Aerospace Computing, Information, and Communication, 3(5), pp. 187–213, 2006.

[2] Masayoshi Matsuoka, Alan Chen, Surya P. N. Singh, Adam Coates, Andrew Y. Ng, Sebastian Thrun. "Autonomous helicopter tracking and localization using a self-surveying camera array." International Journal of Robotics Research., 26(2):205–215, 2007.

[3] A. Bachrach, S. Prentice, R. He, and N. Roy. "RANGE: Robust autonomous navigation in GPS-denied environments". Journal od Field Robotics, 28(5): 644-666, 2011.

[4] M. Acgtelik, A. Bachrach, R. He, S. Prentice, and N. Roy. "Stereo vision and laser odometry for autonomous helicopters in gps-denied indoor environments" in Proceedings of the SPIE Unmanned Systems Technology XI, vol. 7332, Orlando, F, 2009.

[5] Achtelik, M., Roy, N., Bachrach, A., He, R., Prentice, S., and Roy, N., "Autonomous Navigation and Exploration of a Quadrotor Helicopter in GPS-denied Indoor Environments," Proceedings of the 1st Symposium on Indoor Flight, International Aerial Robotics Competition, 2009..

[6] G. Chowdhary, D. M. Sobers Jr., C. Pravitra, C. Christmann, A. Wu, H. Hashimoto, C. Ong, R. Kalghatgi, and E. N. Johnson. "Self-Contained Autonomous Indoor Flight with Ranging Sensor Navigation". Journal of Guidance, Control, and Dynamics, vol. 29, No. 4, November–December. 2012

[7] Bry, A.; Bachrach, A.; Roy, N., "State estimation for aggressive flight in GPS-denied environments using onboard sensing," *Robotics and Automation (ICRA), 2012 IEEE International Conference on* , pp.1-8, 2012.

[8] L. Garcia Carrillo, A. Dzul Lpez, R. Lozano, and C. Pgard, "Combining stereo vision and inertial navigation system for a quad-rotor uav,"Journal of Intelligent & Robotic Systems, pp. 1–15, 2012.

[9] Voigt, Rainer; Nikolic, Janosch; Hurzeler, Christoph; Weiss, Stephan; Kneip, Laurent; Siegwart, Roland; , "Robust embedded egomotion estimation," *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pp.2694-2699, 2011

[10] Weiss, S.; Achtelik, M.W.; Lynen, S.; Chli, M.; Siegwart, R., "Real-time onboard visual-inertial state estimation and self-calibration of MAVs in unknown environments," *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pp.957-964, 2012

[11] Chaolei Wang; Tianmiao Wang; Jianhong Liang; Yang Chen; Yongliang Wu, "Monocular vision and IMU based navigation for a small unmanned helicopter," *Industrial Electronics and Applications (ICIEA), 2012 7th IEEE Conference on*, pp.1694-1699, 2012

[12] Chris Harris and Mike Stephens. "A Combined Corner and Edge Detector". In Alvey Vision Conference, volume 15, page 50. Manchester, UK, 1988.

[13] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 60(2):91-110, 2004.

[14] E. Rosten and T. Drummond. Fusing points and lines for high performance tracking. Computer Vision, 2005. ICCV 2005., 2:1508-1515, 2005

[15] Herbert Bay and Tinne Tuytelaars. SURF: Speeded Up Robust Features. Computer Vision – ECCV 2006, pages 404-417, 2006.

[16] Konolige, K.; Grisetti, G.; Kümmerle, R.; Burgard, W.; Limketkai, B.; Vincent, R.; , "Efficient Sparse Pose Adjustment for 2D mapping," *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on* , vol., no., pp.22-29, 2010

[17] Censi, A.; "An accurate closed-form estimate of ICP's covariance," *Robotics and Automation, 2007 IEEE International Conference on* , vol., no., pp.3167-3172, 2007

[18] A. R. Chowdhury and R. Chellappa, "Stochastic approximation andrate distortion analysis for robust structure and motion estimation,"International Journal of Computer Vision, pp. 27–53, 2003.