

# Cooperative Inchworm Localization with a Low Cost Team

Brian E. Nemsick, Austin D. Buchan, Anusha Nagabandi, Ronald S. Fearing, and Avidesh Zakhor

**Abstract**—In this paper we address the problem of multi-robot localization with a heterogeneous team of low-cost mobile robots. The team consists of a single centralized observer with an inertial measurement unit (IMU) and monocular camera, and multiple picket robots with only IMUs and Red Green Blue (RGB) light emitting diodes (LED). This team cooperatively navigates a visually featureless environment while localizing all robots. A combination of camera imagery captured by the observer and IMU measurements from the pickets and observer are fused to estimate motion of the team. A team movement strategy, referred to as inchworm, is formulated as follows: Pickets move ahead of the observer and then act as temporary landmarks for the observer to follow. This cooperative approach employs a single Extended Kalman Filter (EKF) to localize the entire heterogeneous multi-robot team, using a formulation of the measurement Jacobian to relate the pose of the observer to the poses of the pickets with respect to the global reference frame. An initial experiment with the inchworm strategy has shown localization within 0.14 m position error and 2.18° orientation error over a path-length of 5 meters in an environment with irregular ground, partial occlusions, and a ramp. This demonstrates improvement over a camera-only localization technique that was adapted to our team dynamic which produced 0.18m position error and 3.12° orientation error over the same dataset. In addition, we demonstrate improvement in localization accuracy with an increasing number of picket robots.

## I. INTRODUCTION

The size of a robot can greatly affect what it can do and where it can go. Advantages of small robots include increased accessibility and a wider range of capabilities such as crawling through pipes, inspecting collapsed buildings, exploring congested or complex environments, and hiding in small or inconspicuous spaces. However, these benefits also bring along challenges in the form of reduced sensing abilities, lower communication capability, limited computational resources, and tighter power constraints.

One way to overcome these limitations is to employ a heterogeneous team [1] of collaborative robots. This approach marks a design shift away from the traditional simultaneous localization and mapping (SLAM) ground robots that have expensive sensors and powerful processors, but less mobility in disaster environments. The goal is to have small, mobile, disposable robots with limited capabilities collaborate and share information to accomplish a larger task. Since each robot is expendable, reliability can be obtained in numbers because even if a single robot fails, few capabilities are

This work is supported by the National Science Foundation under the National Robotics Initiative, Award 1427096.

The authors are with Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA {brian.nemsick, abuchan, nagaban2, ronf, avz}@eecs.berkeley.edu

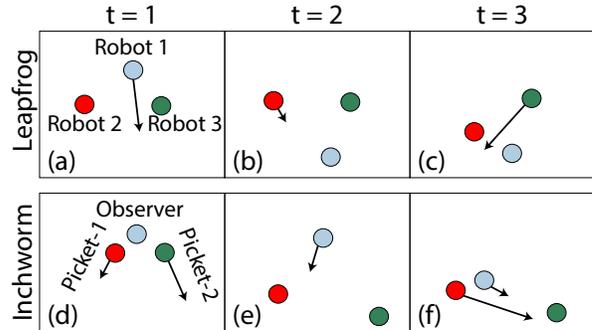


Fig. 1: The above diagrams compare the leapfrog and inchworm strategies. Arrows are drawn to show motion that happens during a time step. In the Leapfrog method (a-c), all robots are the same type and at each time step one robot moves while the other two remain stationary. For example during (a) at  $t = 1$ , robots 2 and 3 remain stationary while robot 1 moves. This process repeats where the moving robot cycles at each time step. In our approach, the inchworm method, at least one robot remains stationary while two move. In addition the picket robots generally remain in front of the observer. For example during (d) the pickets move in front of the observer and during (e) the observer catches up to the stationary pickets. At (f) picket-1 and the observer both move leaving picket-2 stationary.

lost for the team. Hierarchical organization and the idea of a heterogeneous team allows for robots to have different specializations, such as larger robots with higher computation power, smaller robots with increased maneuverability, and robots with different sensor modalities. Another advantage of a team of less capable robots, rather than one extremely capable robot, is that it allows sensing from multiple view-points and hence achieves a wider effective baseline. This is helpful for tasks such as surveillance, exploration, and monitoring. Furthermore, physically traversing an area conveys much more information than simply looking at it from a distance. For example, an expensive scanner can scan the rubble of a disaster site from the outside, but cannot enter and inspect the inside. Knowledge that cannot be gained without physical presence includes detection of slippery surfaces, hidden holes, and other obscured hazards; these can completely incapacitate robots despite their state-of-the-art SLAM algorithms, expensive cameras, and complex laser range finders. Instead, these same hazards can be detected through sacrifice of highly mobile disposable picket robots that scout the area [2].

Localization is a central problem in many robotic applica-

tions. It is particularly important in collaborative situations where without position and orientation information of each robot, there is no global context in which to meaningfully share information between the robots. In this paper, we focus on the localization problem and consider a heterogeneous multi-robot team consisting of two types of minimally equipped robots. A single, central, and more capable observer robot is equipped with a monocular camera and a 6-axis IMU consisting of a gyroscope and accelerometer. Multiple picket robots, which are expendable and less computationally capable, are equipped with no sensors other than 6-axis IMUs. A limited communication interface between the observer robot and individual picket robots is assumed. We consider a multi-robot team with a single observer and multiple picket robots in an unknown environment. We present a method for using a single EKF, which the observer uses to localize the entire multi-robot team, including itself, in six degrees of freedom (6-DOF) by fusing IMU measurements and relative pose estimates of the pickets. Relative pose estimation refers to the process of estimating the position and orientation of a picket's body frame with respect to the camera frame on the observer. This relative pose estimation is done using RGB LEDs that are mounted at known positions on the picket robots.

The datasets discussed in this paper include one observer working together with two pickets to traverse given areas. Even with minimal sensors, the inchworm method is shown to work in dark environments with visual occlusions such as walls or obstacles, instances when line of sight between the robots is lost, and nonplanar settings without external visual features or landmarks. The camera and IMU fusion approach employed by the inchworm method demonstrates improved performance over a camera only approach. In addition, we show that the localization accuracy of the inchworm method improves with an increasing number of picket robots.

## II. RELATED WORK

Existing localization strategies with stationary robots have been explored [1]. A stationary robot is defined as a robot that remains at rest while other robots move. Stationary robots and leapfrogging strategies build on the ideas from [1] and have shown promise in 3-DOF environments in [3] [4]. These previous approaches have a stronger condition than our approach because they require two or three stationary robots at any given time. Our inchworm strategy relaxes these constraints to require only a single stationary robot at any given time, as shown in Figure 1.

A similar approach, cooperative positioning system (CPS), to inchworm is presented in [5]. The CPS approach focuses on 4-DOF ( $x, y, z, \text{yaw}$ ) environments and partitions the robots into two groups, with each group consisting of at least one robot. Under the CPS system, the team alternates between which group moves: Either group A moves or group B moves. For the purposes of comparison we can consider group A to be the observer and group B to be the picket robots. An example of the CPS motion is shown in Figure 1 (d-e) where either the pickets move or the observer moves.

Our inchworm strategy improves on CPS by allowing the observer and picket robots to move at the same time. An example of this is found in Figure 1(f) where both picket-1 and the observer move while picket-2 is stationary. Previous approaches and variants of leapfrogging strategies were focused on team dynamics with high redundancy where each robot produces relative pose estimates of all other robots. Our inchworm approach relaxes the sensor constraints to accommodate teams where only a single observer robot is required to have relative pose estimation capabilities. This leaves the picket robots with more flexibility and less computational burden.

Haldane et al. [2] use a heterogeneous team to detect slippery terrain by sending out a small picket robot and having it walk around the area of interest. The large robot is capable of accurately estimating its own pose, and it uses an augmented reality (AR) tag on the picket robot to localize it. Then, features of the picket's motion are used to train a terrain classifier that is capable of detecting slippery terrain.

A follow-the-leader approach in [6] demonstrates a team composition similar to picket-observer. The leaders and children setup in [7] provides a relative localization scheme in 3-DOF; it assumes accurate localization of the leaders from an external source and localizes the children robots. This approach is extended in [8] to localize the leaders. The problem is subdivided into leader localization and then children localization. The localization of the leaders in [8] requires multiple leaders to maintain line of sight between each other. We extend the approach in [8] to jointly solve the leader and children localization problem without requiring multiple leaders.

A more recent approach [9] uses range-only sensors with a team of aerial vehicles for SLAM and builds on the limited sensor approach of [10]. These drones are equipped with on-board computers and expensive lasers. In contrast, our approach uses inexpensive and disposable picket robots in a 6-DOF environment.

Odometry-based propagation methods have been successful in 3-DOF fusion architectures [1] [11]. However, in 6-DOF non-planar environments, wheel slippage causes systematic biases from encoders. Cell phone quality IMUs are a low cost alternative to wheel encoders in 6-DOF environments because they provide a motion model even under slippage. Extensive work in IMU-based propagation in visual-inertial systems has been explored in [12] [13] [14]. Additionally, monocular pose estimation has been explored in [15] [16].

Many algorithms and approaches exist for multi-robot localization. Graph based approaches have been used [17] [18], and the graph optimization algorithm in [17] relies on the locations of static landmarks and exploits the sparse nature of the graph. Existing EKF [19] [20] [11] or particle filter methods [21] [22] [23] demonstrate the capability of fusing data to provide accurate multi-robot localization.

The noted previous works have extensively and successfully explored multi-robot localization, but their experiments were conducted with access to significantly more capable robots, availability of GPS or beacons of known pose, 3-DOF

---

**Algorithm 1:** Cooperative Inchworm Localization (EKF)

---

**Propagation:** For each IMU measurement:

- buffer previous IMU measurements received from other robots
- propagate state and covariance for the team using the time-step, buffer and new IMU measurement (cf. Section III-B).

**Update:** For each camera image:

- identify RGB LEDs (cf. Section III-C).
- estimate the relative pose between the visible picket robots and the observer frame with P3P and Gauss Newton minimization (cf. Section III-C).
- propagate the state and covariance for the team using the time-step, and most recent IMU measurements
- perform state and covariance update for the team (cf. Section III-D, III-E).

**Inchworm requirement:** At least one stationary robot

---

settings with planar environmental assumptions and accurate wheel odometry, requirements of additional stationary robots, assumptions of light, and existence of landmarks or visual features. In this paper, we relax these assumptions to localize a team consisting of a single observer robot and multiple picket robots. This is accomplished using an EKF approach with the inchworm strategy requirement of at least a single stationary robot at all times. IMU measurements are used for EKF propagation and relative pose estimates are used as an EKF update.

### III. METHODS

The purpose of the multi-robot EKF is to localize all of the robot team's body frames with respect to a global reference frame. An overview of the EKF is provided in Algorithm 1 and can be described as follows: IMU measurements from both types of robots are used to propagate the state and covariance of the team with the same IMU motion model. RGB LEDs are placed with a known configuration on each picket robot such that images captured on the observer can be used to estimate the relative pose of the robots using [24] [25] and Gauss-Newton minimization [15]. Relative pose is defined as the estimation of a picket's body frame, position and orientation, with respect to the camera frame on the observer. The coordinate frames of the team and example LED placement scheme are depicted in Figure 2. The relative pose estimates are subsequently used in the EKF update step.

A team movement strategy called inchworm is adopted, where the picket robots move ahead of the observer to scout and then the observer robot catches up. This movement strategy requires at least one stationary robot. This turn-based approach significantly reduces IMU dead-reckoning error and increases the robustness of the localization algorithm to temporary line of sight. An inchworm increment is a set of motions where the observer and picket robots all move at least once. An example inchworm increment is shown in Figure 1 (d-e). A stationary robot does not propagate

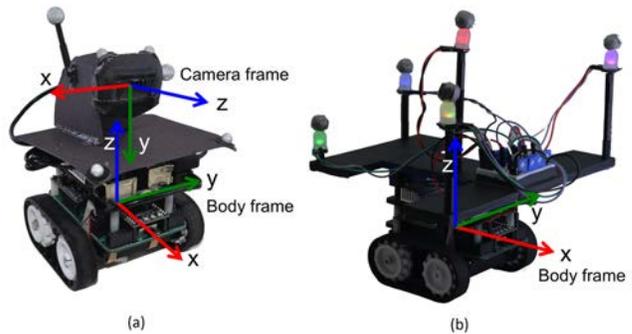


Fig. 2: Coordinate frame overview for a sample team consisting of two robots. The observer, (a), is mounted with a camera and the picket, (b), with multi-color LED markers.

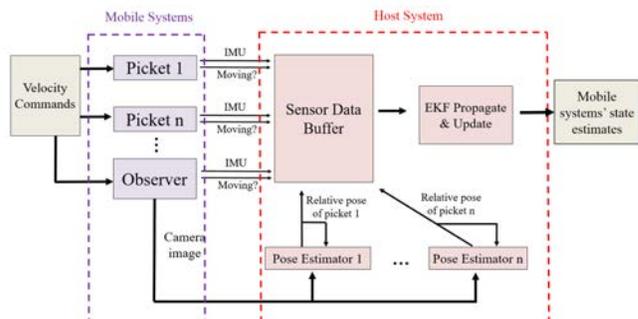


Fig. 3: Block diagram of the asynchronous multi-robot team performing real-time cooperative localization algorithm. Asynchronous sensor data from the robots is sent over WiFi, sorted into a measurement buffer, and then used in the EKF propagate and update step. Currently, the host system is an external laptop.

its corresponding states or covariances, thus bounding the uncertainty of the entire team. This enables the stationary robot to function as a temporary visual landmark and serves as a functional substitute to external visual features. Although external visual features are used in traditional visual odometry or visual SLAM systems, they are not consistently available in low light environments.

One benefit of a stationary picket robot is in situations of complete line of sight failure, where none of the picket robots are visible to the observer. In this case, a single future re-observation of a stationary robot, i.e. loop-closure, corrects the IMU dead-reckoning error of the non-stationary robots.

The following sections describe the EKF propagation and update steps in detail.

#### A. State Vector

The EKF state and accompanying error-state vector stores the state of each single-robot in the multi-robot team. The state vector components with respect to  $i^{\text{th}}$  picket robot are:

$$\mathbf{x}_i = [{}^B_G\mathbf{q}_i^T, {}^G\mathbf{p}_i^T, {}^G\mathbf{v}_i^T, \mathbf{b}_{i_g}^T, \mathbf{b}_{i_a}^T]^T \in \mathbb{R}^{16 \times 1} \quad (1)$$

where  ${}^B_G\mathbf{q}_i^T \in \mathbb{R}^{4 \times 1}$ , is the unit quaternion representation of the rotation from the global frame  $\{G\}$  to the body frame

$\{\mathbf{B}\}$ ,  ${}^G\mathbf{p}_i, {}^G\mathbf{v}_i \in \mathbb{R}^{3 \times 1}$  are the body frame position and velocity with respect to the global frame, and  $\mathbf{b}_{i_g}, \mathbf{b}_{i_a} \in \mathbb{R}^{3 \times 1}$  are the gyroscope and accelerometer biases.

The corresponding error-state components with respect to  $i^{\text{th}}$  picket robot are:

$$\tilde{\mathbf{x}}_i = [{}^G\tilde{\boldsymbol{\theta}}_i^T, {}^G\tilde{\mathbf{p}}_i^T, {}^G\tilde{\mathbf{v}}_i^T, \tilde{\mathbf{b}}_{i_g}^T, \tilde{\mathbf{b}}_{i_a}^T]^T \in \mathbb{R}^{15 \times 1} \quad (2)$$

where  ${}^G\tilde{\boldsymbol{\theta}}_i^T$  is the minimal representation from the error quaternion  $\delta\tilde{\mathbf{q}} \simeq [\frac{1}{2}{}^G\tilde{\boldsymbol{\theta}}_i^T, 1]^T$  [13] [14]. The non-quaternion states use the standard additive error model.

The observer robot is also a component in the EKF state and error-state vector:

$$\begin{aligned} \mathbf{x}_o &= [{}^O\tilde{\mathbf{q}}_o^T, {}^G\tilde{\mathbf{p}}_o^T, {}^G\tilde{\mathbf{v}}_o^T, \mathbf{b}_{o_g}^T, \mathbf{b}_{o_a}^T]^T \in \mathbb{R}^{16 \times 1} \\ \tilde{\mathbf{x}}_o &= [{}^G\tilde{\boldsymbol{\theta}}_o^T, {}^G\tilde{\mathbf{p}}_o^T, {}^G\tilde{\mathbf{v}}_o^T, \tilde{\mathbf{b}}_{o_g}^T, \tilde{\mathbf{b}}_{o_a}^T]^T \in \mathbb{R}^{15 \times 1} \end{aligned} \quad (3)$$

where  $\{O\}$  denotes the observer frame.

Combining the states in Eqns. 1, 2, and 3, the augmented EKF state vector and error-state vector with respect to the multi-robot team with  $n$  pickets becomes:

$$\begin{aligned} \mathbf{x} &= [\mathbf{x}_o^T, \mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_n^T]^T \in \mathbb{R}^{16(n+1) \times 1} \\ \tilde{\mathbf{x}} &= [\tilde{\mathbf{x}}_o^T, \tilde{\mathbf{x}}_1^T, \tilde{\mathbf{x}}_2^T, \dots, \tilde{\mathbf{x}}_n^T]^T \in \mathbb{R}^{15(n+1) \times 1} \end{aligned} \quad (4)$$

where  $n$  is the total number of picket robots.

### B. IMU Propagation Model

The EKF propagation step occurs each time a new IMU measurement from any single-robot or a camera image is captured on the observer robot. The continuous dynamics of the IMU propagation model for a single-robot are [13] [14]:

$$\begin{aligned} {}^B\dot{\mathbf{q}} &= \frac{1}{2}\boldsymbol{\Omega}(\boldsymbol{\omega})_G^B\tilde{\mathbf{q}}, \quad {}^G\dot{\mathbf{p}} = {}^G\mathbf{v}, \quad {}^G\dot{\mathbf{v}} = {}^G\mathbf{a} \\ \dot{\mathbf{b}}_g &= \mathbf{n}_{wg}, \quad \dot{\mathbf{b}}_a = \mathbf{n}_{wa} \end{aligned} \quad (5)$$

where  $\mathbf{n}_{wg}, \mathbf{n}_{wa}$  are Gaussian white noise vectors for the gyroscope and accelerometer respectively and

$$[\boldsymbol{\omega} \times] = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}, \quad \boldsymbol{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} -[\boldsymbol{\omega} \times] & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^T & 0 \end{bmatrix} \quad (6)$$

The discrete time linearized model and the error-state model are derived and discussed with detail in [13] [14].

Critically, stationary robots receive no state or covariance propagation. This prevents IMU dead-reckoning drift from moving a temporary landmark and maintains a bounded covariance block pertaining to the stationary robot.

### C. Relative Pose Estimation

Four (or more) RGB LEDs are placed at known configurations position on the picket robots to allow relative pose estimation on board the observer. Each picket robot receives a unique configuration with LEDs of various colors. Color and intensity thresholds are used to find the LED centroids, and these LED detections are passed into separate pose estimators (one pose estimator for each robot).

From the centroid detections, the approach from [15] is used to perform relative pose estimation. Pose correspondence is computed with the perspective-3-point (P3P) [24]

[25] algorithm for each picket. Using different colors for the LEDs reduces the computational load by allowing the P3P correspondence search to search fewer possible configurations. Gauss-Newton minimization refines the initial solution from the P3P algorithm by minimizing reprojection error [15]:

$$P^* = \arg \min_P \sum_{\langle \mathbf{l}, \mathbf{d} \rangle \in \mathbf{C}} \|\pi(\mathbf{l}, P - \mathbf{d})\|^2$$

where  $P$  is pose estimate,  $\mathbf{l}$  is the set of LED configurations,  $\mathbf{d}$  is the set of LED centroids,  $\mathbf{C}$  is the LED correspondences, and  $\pi$  projects an LED from  $\mathbb{R}^3$  into  $\mathbb{R}^2$  (camera image).

The pose estimate covariance ( $\mathbf{Q}$ ) is calculated with the Jacobian ( $\mathbf{J}$ ) from the Gauss-Newton minimization [15]:

$$\mathbf{Q} = (\mathbf{J}^T \boldsymbol{\Sigma}^{-1} \mathbf{J})^{-1} \text{ where } \boldsymbol{\Sigma} = \mathbf{I}_{2 \times 2} \text{ pixels}^2 \quad (7)$$

### D. Camera Measurement Model

In this section we describe how the relative pose estimates are used to compute the EKF update. We derive the residual and observation matrix that relates the relative pose estimates to the state vector as described in Section III-A. The residual and the observation matrix are used to calculate the Kalman gain and correction.

An example overview of the multi-robot teams coordinates frames is shown in Figure 2. A static camera transform is defined as:

$$[{}^C\tilde{\mathbf{q}}^T, {}^C\tilde{\mathbf{p}}^T]^T \in \mathbb{R}^{7 \times 1} \quad (8)$$

With respect to a single visible picket robot,  $i$ , a relative pose estimate from the camera frame on board the observer is defined as:

$$\mathbf{z}_i = [{}^B\tilde{\mathbf{q}}_{i_z}^T, {}^B\tilde{\mathbf{p}}_{i_z}^T]^T \in \mathbb{R}^{7 \times 1} \quad (9)$$

In an EKF framework, a residual,  $\mathbf{r}$ , and a measurement Jacobian,  $\mathbf{H}$  are used to compute the EKF update. The standard relationship between the residual and measurement Jacobian is:

$$\mathbf{r} = \mathbf{z} - \hat{\mathbf{z}} \approx \mathbf{H}\tilde{\mathbf{x}} + \mathbf{n} \quad (10)$$

where  $\mathbf{n}$  is noise. A prediction of the observation,  $\hat{\mathbf{z}}_i$ , is used to compute a residual in an EKF. This observation corresponds to a relative pose for each visible robot. Additionally, the quaternion states in  $\mathbf{x}$  use the rotational error definition,  $\delta\mathbf{q} = \mathbf{q} \otimes \hat{\mathbf{q}}^{-1}$  rather than the standard linear error,  $\tilde{\mathbf{p}} = \mathbf{p} - \hat{\mathbf{p}}$ .

To compute  $\hat{\mathbf{z}}_i$ , the state vector estimate is updated with the EKF propagation step. The poses of the picket robots are then converted from the global frame converted to the camera coordinate frame,  $\{C\}$ , in Eq. 8 to match the relative pose estimate:

$$\hat{\mathbf{z}}_i = \begin{bmatrix} {}^B\hat{\mathbf{q}}_i \\ {}^B\hat{\mathbf{p}}_i \end{bmatrix} = \begin{bmatrix} {}^B\hat{\mathbf{q}}_i \otimes {}^G\hat{\mathbf{q}}_O \otimes {}^C\hat{\mathbf{q}}_O \\ {}^C\mathbf{R}_O^B \mathbf{R}_i^G ({}^B\hat{\mathbf{p}}_i - {}^G\hat{\mathbf{p}}_O - {}^C\hat{\mathbf{p}}) \end{bmatrix} \quad (11)$$

where  $\otimes$  represents quaternion multiplication.

The single-robot residuals with respect to each visible pickets robots are calculated according to the definition Eq. 10:

$$\mathbf{r}_i = \mathbf{z}_i - \hat{\mathbf{z}}_i = \begin{bmatrix} 2 \cdot \pi \left( \begin{matrix} {}^B \hat{\mathbf{q}}_i^{-1} \otimes {}^B \tilde{\mathbf{q}}_{i_z} \\ {}^B \tilde{\mathbf{p}}_{i_z} - {}^B \tilde{\mathbf{p}}_i \end{matrix} \right) \end{bmatrix} \quad (12)$$

where  $\pi$  is defined as  $\pi([\mathbf{q}_x, \mathbf{q}_y, \mathbf{q}_z, \mathbf{q}_w]^T)^T = [\mathbf{q}_x, \mathbf{q}_y, \mathbf{q}_z]^T$  and utilized as a small angle approximation for the orientation difference between  $\mathbf{z}_i$  and  $\hat{\mathbf{z}}_i$ .

The  $i^{\text{th}}$  measurement Jacobian,  $\mathbf{H}_i$ , is calculated by applying small angle approximations and taking the partial derivatives of the  $i^{\text{th}}$  single-robot residual with respect to the error-state. The non-zero entries are shown below:

$$\begin{aligned} \mathbf{r}_i &\simeq \mathbf{H}_i \tilde{\mathbf{x}} \\ \mathbf{H}_i &= \begin{bmatrix} {}^G \hat{\mathbf{R}} \mathbf{L} ({}^B \tilde{\mathbf{p}}_i - {}^G \tilde{\mathbf{p}}_o - {}^G \tilde{\mathbf{p}} \times \mathbf{J}) & -{}^G \hat{\mathbf{R}} & \mathbf{0} & \dots & {}^G \hat{\mathbf{R}} & \mathbf{0} & \mathbf{0} & \dots \\ -{}^G \hat{\mathbf{R}} & \mathbf{0} & \dots & {}^G \hat{\mathbf{R}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots \end{bmatrix} \\ \tilde{\mathbf{x}} &= [{}^G \tilde{\boldsymbol{\theta}}_o \quad {}^G \tilde{\mathbf{p}}_o \quad {}^G \tilde{\mathbf{v}}_o \quad \dots \quad {}^G \tilde{\boldsymbol{\theta}}_i \quad {}^G \tilde{\mathbf{p}}_i \quad {}^G \tilde{\mathbf{v}}_i \quad \dots] \end{aligned} \quad (13)$$

where  ${}^G \hat{\mathbf{R}} = {}^G \mathbf{R}_O^G \hat{\mathbf{R}}$  and  $[\mathbf{q} \times]$  is the quaternion skew operator from Eq. 6. The higher order and cross terms are dropped from  $\mathbf{H}_i$  to satisfy the linear requirement of the EKF.

The states of all picket robots become correlated with the observer robot through the measurement Jacobian. This enables an individual pose estimate of a picket robot to improve the state estimate of each picket robot. The correlation is essential to localizing the observer robot because it is unable to observe itself directly from camera imagery.

#### E. EKF Update

From the camera measurement model the EKF update is performed. To utilize the standard equations, the overall measurement Jacobian is calculated by vertically stacking the single-robot measurement Jacobians from the camera measurement model in Eq. 13:

$$\mathbf{H} = [\mathbf{H}_1^T, \mathbf{H}_2^T, \dots, \mathbf{H}_n^T]^T \in \mathbb{R}^{6n \times 16(n+1)} \quad (14)$$

Accordingly the measurements,  $\mathbf{z}_i$ , are stacked identically:

$$\mathbf{z} = [\mathbf{z}_1^T, \mathbf{z}_2^T, \dots, \mathbf{z}_n^T]^T \in \mathbb{R}^{7n \times 1} \quad (15)$$

The corresponding overall observation noise is calculated by diagonalizing the uncorrelated relative pose estimate covariances from Eq. 7:

$$\mathbf{Q} = \text{diag}(\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_n) \in \mathbb{R}^{6n \times 6n} \quad (16)$$

From Eqs. 14, 15, and 16, the procedure to update an EKF with quaternion states is described in [13] [14].

## IV. RESULTS

### A. Experimental Approach

We apply the localization technique described above to data collected from a team of three small, low-cost, mobile robots. The Zummy robot<sup>1</sup> is a decimeter-scale tracked robot running ROS on board a Linux computing system with networking and vision processing capabilities. The observer Zummy supports a Microsoft Lifecam 3000 webcam with

<sup>1</sup><https://wiki.eecs.berkeley.edu/biomimetics/Main/Zummy>

$640 \times 480$  pixels<sup>2</sup> at 30 Hz, InvenSense MPU-6050 MEMS IMU at 30 Hz, and supports WiFi wireless communication. This robot is designed to be easily built from commercially available off-the-shelf parts for a total cost of  $\approx$  \$350.

The robotic team consists of one observer and two picket robots shown in Figure 2. A Zummy robot with a camera serves as the observer, and to represent the inexpensive and less capable picket robots, we use Zummy robots without cameras. Each picket robot is outfitted with an LED “hat” so that it can be visually tracked by the observer robot. Infrared markers are also attached to each Zummy in order to obtain ground truth from a VICON motion capture system. The robots are manually driven for these datasets.

### B. Planar Base Case

The baseline experimental task was a cooperative U-turn in planar 3-DOF with one observer and two pickets. The robots were manually driven in the dark. Although the dataset was recorded in a 3-DOF environment, the filter was not constrained with environmental priors. A direct comparison between the LED pose tracker system setup in [15] and our system setup is in Table I.

TABLE I: Pose Tracker Comparison

	Faessler et al. [15]	Our System
Resolution (pixels <sup>2</sup> )	752x480	640x480
Baseline Radius (cm)	10.9	10.6
LEDs/Robot	5	5
LED Type	Infrared	RGB
$\approx$ error at 2 m depth	5 cm, 1-2 deg	5-8 cm, 1-4 deg

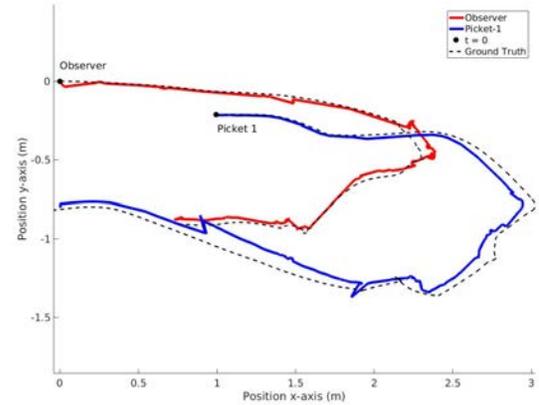


Fig. 4: A plot of the XY projection of the team’s pose estimates from the EKF along with the ground truth trajectories. Shown for the base case of the planar U-turn where a single picket is used to perform localization.

In Figures 4 and 5, we show the resulting trajectories from the localization of all team members during this U-turn dataset and we compare to ground truth. We plot the results of using only one picket while discarding the measurements from other, and then the results of using both pickets. Note that the observer trajectory is not as smooth as that of

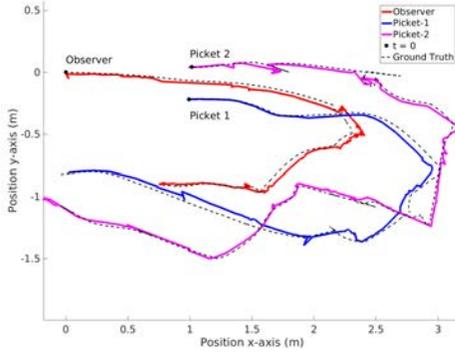


Fig. 5: A plot of the XY projection of the team’s pose estimates from the EKF along with the ground truth trajectories. Shown for the base case of the planar U-turn where both pickets are used to perform localization.

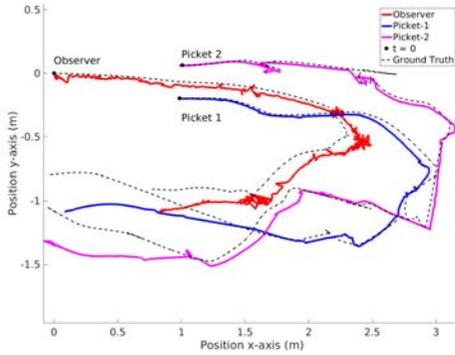


Fig. 6: Camera-only approach: A plot of the XY projection of the team’s pose estimates along with the ground truth trajectories, using a camera-only approach. Performs notably worse in yaw drift than the IMU-camera fusion approach shown in Figures 4, 5.

TABLE II: Planar Drift Analysis

		Camera-only Two Pickets	Fusion One Picket	Fusion Two Pickets
Observer	x (cm)	-8.14	<b>0.67</b>	-0.80
	y (cm)	16.64	2.58	<b>-1.46</b>
	z (cm)	4.09	-5.42	<b>3.07</b>
	Angle (°)	9.33	1.89	<b>1.54</b>
Picket-1	x (cm)	-13.38	<b>-4.25</b>	-6.23
	y (cm)	28.97	-1.79	<b>-1.70</b>
	z (cm)	4.53	-7.91	<b>4.39</b>
	Angle (°)	4.72	2.64	<b>1.36</b>
Picket-2	x (cm)	<b>-1.35</b>	-	9.92
	y (cm)	25.40	-	<b>-5.65</b>
	z (cm)	4.56	-	<b>5.18</b>
	Angle (°)	4.56	-	<b>1.88</b>

picket-1 or picket-2, because the motion of the observer has un-modeled vibration effects that cause motion blur and temporary changes to the “static” camera transform.

This dataset consisted of 10 inchworm increments. An inchworm increment is defined as the minimal set of team motions where each robots has moved once. End-position

drift for using a single picket vs. using two pickets are shown in Table II. The angular drift, which causes propagates error into the future, for the two picket fusion case was less than the one picket case. Although unconstrained to a plane, the angular drift was almost exclusively in yaw. Performing right or left turns with the robot team introduces more rotational drift than forward or backwards motions. Without external features or global correction, the yaw errors persist until the end of experiment, but adding more picket robots helps to mitigate these effects. The jagged regions of the trajectory correspond to the observer and picket robots starting or stopping motion, and they are due to The LED mounts and the robots shaking during these transient motions.

A camera-only filtering approach was evaluated in Figure 6 as a baseline. The camera-only approach uses the same formulation of the measurement derived in Section III-D but without a motion model. This method performed significantly worse with four times as much yaw drift than the IMU plus camera fusion approach. Without the gyroscope, the inchworm localization performs significantly worse in orientation estimation.

### C. Non-planar Terrain with Ramp

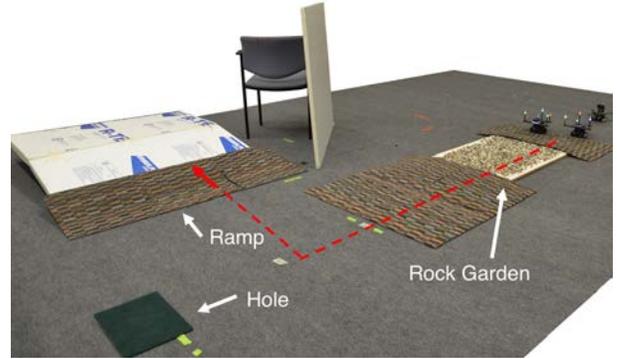


Fig. 7: 6-DOF environment for testing: Robot team is on the right, rock garden is center-right), a “hole” is shown on the bottom-left, and the ramp is on top left.

The second experiment was conducted in an environment featuring non-planar terrain, obstacles, and occlusions. The robots were manually driven in the environment shown in Figure 7. The 6-DOF non-planar dataset consisted of 10 inchworm increments: 3 for the rock garden and 7 for the right turn and ramp. Temporary line of sight failure of both pickets occurred during the rock garden because the pose estimator failed to converge, as the observer was moving on the rocks. Wheel slippage also occurred during the rock garden section. After the rock garden, picket-2 was deliberately left behind to simulate a hole in the environment and a loss of a robot.

The ground truth trajectories and the EKF pose estimates of the dataset are shown in Figures 9 and 10. The end point drift analysis is shown in Table III with a comparison against a camera-only approach. The fusion approach outperformed the camera-only for the observer and picket-1. The most

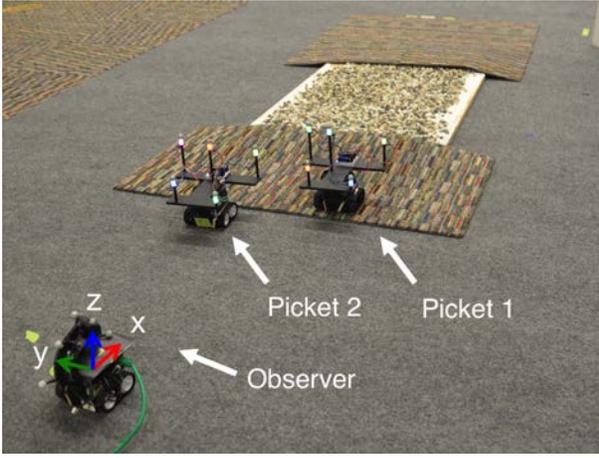


Fig. 8: Starting position of the robot team with view of the rock garden section. The origin is defined as the starting position of the observer.

critical improvement is the orientation error of the observer, which persists without correction. Picket-2 traveled mostly in a straight line except during the rock garden, and the endpoint errors of both approaches are almost identical. The drift is predominantly in pitch for each robot. Temporal plots with ground truth are in Figures 11 and 12.

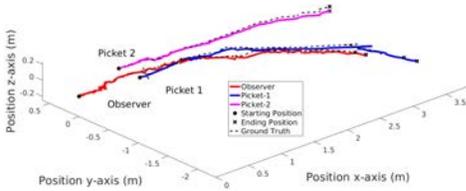


Fig. 9: Ground truth trajectories of the multi-robot team are compared against the estimates of the EKF for the non-planar environment. Axes are scaled equally

TABLE III: Non-Planar Drift Analysis

		Camera-only Two Pickets	Fusion Two Pickets
Observer	x (cm)	<b>-4.03</b>	-5.35
	y (cm)	13.36	<b>12.67</b>
	z (cm)	1.01	<b>0.04</b>
	Angle (°)	3.12	<b>2.18</b>
Picket-1	x (cm)	<b>-2.11</b>	-4.48
	y (cm)	17.9	<b>16.76</b>
	z (cm)	1.72	<b>0.10</b>
	Angle (°)	6.22	<b>4.29</b>
Picket-2	x (cm)	<b>0.28</b>	0.32
	y (cm)	-0.20	<b>0.15</b>
	z (cm)	<b>5.37</b>	5.37
	Angle (°)	<b>3.58</b>	3.59

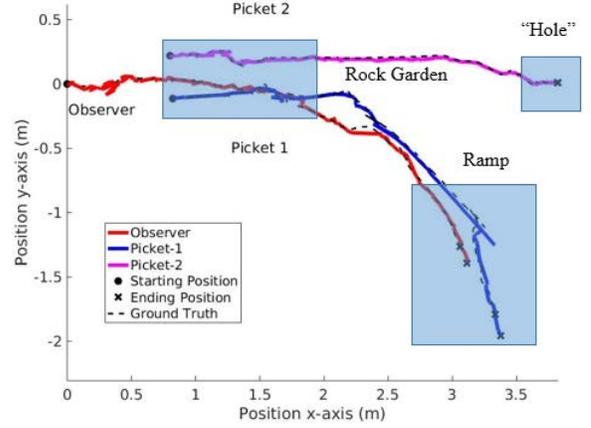


Fig. 10: 2D projection of ground truth trajectories of the multi-robot team are compared against the estimates of the EKF for the non-planar environment.

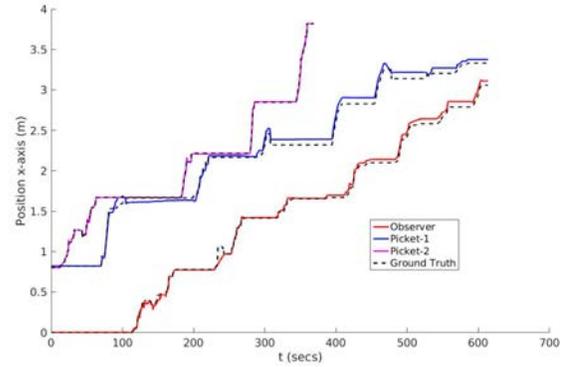


Fig. 11: Position along the x-axis versus time.

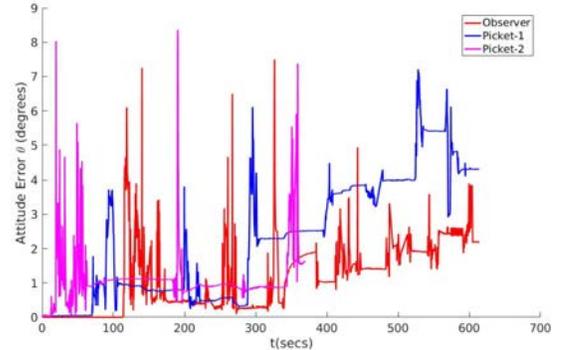


Fig. 12: Orientation error versus time.

## V. CONCLUSION

A heterogeneous team which consists of a single observer and multiple picket robots is able to navigate a visually featureless, unknown, non-planar environment as a unit, using only relative pose observations and IMU measurements to estimate the motion of the entire team. The IMU and camera fusion approach presented in this paper has clear

advantages over a simpler camera-only approach (Figures 4, 5, and 6) and its benefits outweigh the cost of having asynchronous communication. Although calibration of an IMU adds many complications, it is a natural choice for environments in which wheel encoders are unreliable. A camera-only approach heavily relies on line of sight at all times, which is restricting and potentially impractical to maintain. In addition, motion-model based approaches for EKF propagation allow for the rejection of errant camera pose estimates from faulty LED detections or P3P correspondence matching. Most importantly, with a camera-only approach, each inchworm increment has an associated positional and rotational drift in 6-DOF. The calibration of the IMU allows the fusion based approach of using the stationary robots' gravity vectors to reduce and bound the pitch and roll drift leading to drift in only 4-DOF.

In the future, we will create exploration strategies for larger robot teams of more than 10 robots. With this increasing number of robots, autonomous control is far more effective than manual driving. An advanced control scheme that factors in terrain, obstacles, and collaboration of robots will be developed for effective exploration in hazardous environments.

#### REFERENCES

- [1] R. Grabowski, L. E. Navarro-Serment, C. J. Paredis, and P. K. Khosla, "Heterogeneous teams of modular robots for mapping and exploration," *Autonomous Robots*, vol. 8, no. 3, pp. 293–308, 2000.
- [2] D. W. Haldane, P. Fankhauser, R. Siegwart, and R. S. Fearing, "Detection of slippery terrain with a heterogeneous team of legged robots," in *IEEE Int. Conf. on Robotics and Automation*, 2014, pp. 4576–4581.
- [3] L. E. Navarro-Serment, C. J. Paredis, P. K. Khosla, *et al.*, "A beacon system for the localization of distributed robotic teams," in *Int. Conf. on Field and Service Robotics*, vol. 6, 1999.
- [4] S. Tully, G. Kantor, and H. Choset, "Leap-frog path design for multi-robot cooperative localization," in *Int. Conf. on Field and Service Robotics*. Springer, 2010, pp. 307–317.
- [5] R. Kurazume and S. Hirose, "An experimental study of a cooperative positioning system," *Autonomous Robots*, vol. 8, no. 1, pp. 43–52, 2000.
- [6] K. E. Wenzel, A. Masselli, and A. Zell, "Visual tracking and following of a quadcopter by another quadcopter," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 4993–4998.
- [7] T. R. Wanasinghe, G. K. Mann, and R. G. Gosine, "Distributed collaborative localization for a heterogeneous multi-robot system," in *27th IEEE Canadian Conf. on Electrical and Computer Engineering (CCECE)*, 2014, pp. 1–6.
- [8] —, "Distributed leader-assistive localization method for a heterogeneous multirobotic system," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 3, pp. 795–809, 2015.
- [9] F. Fabresse, F. Caballero, and A. Ollero, "Decentralized simultaneous localization and mapping for multiple aerial vehicles using range-only sensors," in *IEEE Int. Conf. on Robotics and Automation*, 2015, pp. 6408–6414.
- [10] K. E. Bekris, M. Glick, and L. E. Kavraki, "Evaluation of algorithms for bearing-only slam," in *IEEE Int. Conf. on Robotics and Automation*, 2006, pp. 1937–1943.
- [11] R. Madhavan, K. Fregene, and L. E. Parker, "Distributed heterogeneous outdoor multi-robot localization," in *IEEE Int. Conf. on Robotics and Automation*, 2002, vol. 1, pp. 374–381.
- [12] D. Strelow and S. Singh, "Motion estimation from image and inertial measurements," *The International Journal of Robotics Research*, vol. 23, no. 12, pp. 1157–1195, 2004.
- [13] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *IEEE Int. Conf. on Robotics and Automation*, 2007, pp. 3565–3572.
- [14] G. Huang, M. Kaess, and J. J. Leonard, "Towards consistent visual-inertial navigation," in *IEEE Int. Conf. on Robotics and Automation*, 2014, pp. 4926–4933.
- [15] M. Faessler, E. Mueggler, K. Schwabe, and D. Scaramuzza, "A monocular pose estimation system based on infrared LEDs," in *IEEE Int. Conf. on Robotics and Automation*, 2014, pp. 907–913.
- [16] A. Breitenmoser, L. Kneip, and R. Siegwart, "A monocular vision-based system for 6D relative robot localization," in *IEEE Int. Conf. on Intelligent Robots and Systems*, 2011, pp. 79–85.
- [17] A. Ahmad, G. D. Tipaldi, P. Lima, and W. Burgard, "Cooperative robot localization and target tracking based on least squares minimization," in *IEEE Int. Conf. on Robotics and Automation*, 2013, pp. 5696–5701.
- [18] V. Indelman, E. Nelson, N. Michael, and F. Dellaert, "Multi-robot pose graph localization and data association from unknown initial relative poses via expectation maximization," in *IEEE Int. Conf. on Robotics and Automation*, 2014, pp. 593–600.
- [19] S. I. Roumeliotis and G. A. Bekey, "Distributed multirobot localization," *IEEE Transactions on Robotics and Automation*, vol. 18, no. 5, pp. 781–795, 2002.
- [20] A. Martinelli, F. Pont, and R. Siegwart, "Multi-robot localization using relative observations," in *IEEE Int. Conf. on Robotics and Automation*, 2005, pp. 2797–2802.
- [21] A. Howard, "Multi-robot simultaneous localization and mapping using particle filters," *The International Journal of Robotics Research*, vol. 25, no. 12, pp. 1243–1256, 2006.
- [22] L. Carlone, M. K. Ng, J. Du, B. Bona, and M. Indri, "Rao-Blackwellized particle filters multi robot SLAM with unknown initial correspondences and limited communication," in *IEEE Int. Conf. on Robotics and Automation*, 2010, pp. 243–249.
- [23] A. Prorok, A. Bahr, and A. Martinoli, "Low-cost collaborative localization for large-scale multi-robot systems," in *IEEE Int. Conf. on Robotics and Automation*, 2012, pp. 4236–4241.
- [24] L. Quan and Z. Lan, "Linear n-point camera pose determination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 774–780, 1999.
- [25] L. Kneip, D. Scaramuzza, and R. Siegwart, "A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 2969–2976.