Multi-modal Sensor Fusion Algorithm for Ubiquitous Infrastructure-free Localization in Vision-impaired Environments

Taragay Oskiper, Han-Pang Chiu, Zhiwei Zhu, Supun Samarasekera, Rakesh Kumar

***Abstract*— In this paper, we present a unified approach for a camera tracking system based on an error-state Kalman filter algorithm using both relative (local) measurements obtained from image based motion estimation through visual odometry, and global measurements as a result of landmark matching through a pre-built visual landmark database and range measurements obtained from radio frequency (RF) ranging radios. We show our results by using the camera poses output by our system to render views from a 3D graphical model built upon the same frame as the landmark database which also forms the global coordinate system and compare them to the actual video images. These results help demonstrate both the long term stability and the overall accuracy of our algorithm as intended to provide a solution to the GPS denied ubiquitous camera tracking problem under both vision-aided and vision-impaired conditions.**

I. Introduction

In this paper, we present our recent work on a real time navigation system, which can be used both indoors and outdoors over large areas in GPS challenged environments. The navigation and localization module consists of EO stereo sensors, micro-electro-mechanical type inertial measurement unit (MEMS IMU), ranging radios, and a COTS PC based processor package. The navigation system can be used for many applications including multi-robot human coordination, multi-robot control and augmented reality for humans wearing optical see-through HMDs. In each case the 3D position and orientation of the robot or human is tracked by processing the data from robot/human-worn optical, IMU and radio frequency (RF) sensors. Exploiting the multiple-sensor data provides several layers of robustness to the navigation system built upon a visual odometry and visual landmark matching based backbone. Sensor data from a low cost MEMS IMU and RF ranging between mobile humans/robots and static RF nodes is fused with the vision information by a Kalman filter to provide robustness under challenging conditions where there are insufficient visual clues to rely on. Furthermore the fusion of vision aided navigation module with RF-ranging between robots/humans enables improved localization of each entity in a single coordinate system.

The mobile localization problem has been extensively studied in robotics community. Mobile robot localization is the problem of estimating the position and orientation of a robot relative to its environment. According to [8], there are three classes of sub-problems in this area: position tracking, global localization, and kidnapped robot. By giving the odometry incremental measurements and other information such as range, a localization algorithm tries to estimate the position of a robot. If the incremental measurement error is small and the kinematical and measurement noise satisfy the Gaussian assumption, the Kalman filter algorithm is a good candidate to solve the problem [2],[7]. In our approach, we use the distributed aperture visual odometry algorithm with front and back facing stereo cameras and MEMS IMU as described in [4] and the visual landmark matching framework introduced in [9] and track 6 DOF pose to centimeter level accuracy. The main differences in this paper are in that, we use a new IMU-centric Kalman filter framework eliminating the constant velocity process model used in [4] and instead of the final system pose being generated by the landmark matching module as was done in [9], the 3D to 2D point correspondences between the database and given query image are input as global measurements into the filter. This unified approach to fuse all the visual measurement data allows for better handling of the uncertainty propagation through the

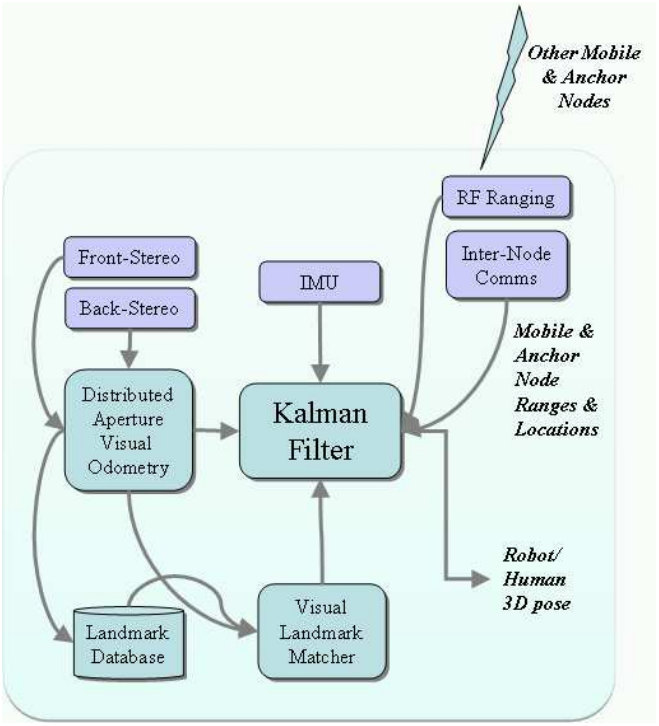

Fig. 1. System block diagram.

Taragay Oskiper, Han-Pang Chiu, Zhiwei Zhu are with the Vision and Robotics Laboratory, Sarnoff Corporation, 201 Washington Road, Princeton, NJ 08540, USA {toskiper,hchiu,zzhu}@sarnoff.com

Supun Samarasekera, Rakesh Kumar are with the Vision and Robotics Laboratory, Sarnoff Corporation, 201 Washington Road, Princeton, NJ 08540, USA {ssamarasekera,rkumar}@sarnoff.com

whole system, not possible in our earlier framework in which the Kalman filter output was used to locally propagate the navigation solution from one landmark match instance to another such that the pose solution obtained as a result of landmark matching would effectively reset the filter output. Furthermore, the current Kalman filter is also provided with RF range measurements to enable navigation under vision-impaired environments (cf. Fig. 1). For this purpose, we have employed the Nanotron range measuring radio as our range sensor, which uses short (3ns) pulses converted to chirps (frequency sweeps), and has very good noise and multipath immunity. It has a working range of greater than 150m outdoors, and it offers a compact solution (35mm x 14mm x 3mm) with low power (500mW) and low cost making it very attractive. In our application, we have used five of these ranging radios, two of which are deployed as static anchor nodes at known locations in the exercise area. (The coordinates of these locations are known and obtained from the graphical model built on our landmark database.) The other three radios are used as mobile nodes that are mounted on the backpack units carried by the users with the radio antenna attached next to the front left camera on the helmets, this being the master camera that is tracked in all the systems. The locations of these units are broadcast over the wireless network at frame-rate (15Hz) across all the users. For this purpose we use an open source data distribution service approach where each user publishes its own camera location packets as well as subscribes to all the incoming location information sent by the others. Radio packets carrying range data are associated with camera location packets received by each unit based on the ID field in each packet and synchronized inside a preprocessing module before they are fed into the Kalman filter as measurements.

In order to provide global fixes to prevent the camera poses from drifting during online tracking, a landmark database of the area where the exercise will take place is created before-hand. Mainly, a pan-tilt unit captures both Lidar and stereo imagery while panning full 360 degrees at regularly spaced intervals. All the data is processed offline in a semi-automated manner to produce high fidelity camera poses for each landmark shot stored in the database that includes the image feature coordinates together with their locally triangulated 3D coordinates (expressed in the left camera frame), the reconstruction uncertainty of each point represented by a covariance matrix and the associated camera pose for that shot. Also feature descriptors are entered into a vocabulary tree to allow fast indexing during online exercise [9].

We adopt the so called "error-state" formulation for our extended Kalman filter for several reasons [6]. Under this representation, there is no need to specify an explicit dynamic motion model for a given sensor platform since the IMU captures with high fidelity the short term high frequency motion of the rig as represented by the mechanization equations (4)-(6). Thus the same model can be used whether the sensor rig is placed on a robot or on a human-worn helmet as in our case. The filter dynamics follow from the IMU error propagation equations which evolve slowly over time and therefore are more amenable to linearization. The measurements to the filter consist of the differences between the inertial navigation solution as obtained by solving the IMU mechanization equations and the external source data, which in our case is the relative pose information provided by visual odometry algorithm, global measurements provided by the visual landmark matching process, and ranges to the radio nodes in the environment (either static or mobile) which constitute global measurements when combined with their locations.

Note that, in this sensor configuration, the range measurements from RF sensors play a minimal role during the periods when the vision based sensors are performing well, due to the fact that the global positional uncertainty of the system remains below the radio measurement noise level when visual odometry and landmark matching modules are functioning. However, during those periods when the vision based solution is failing, such as when the user enters into thick smoke as in our experiments (cf. supplemental video material) or into a low light environment, the positional uncertainty increases rapidly due to the nature of unaided IMU based navigation, to the level where the system seamlessly transitions into relying more on the range measurements which keeps the tracking under check and maintains drift-free navigation. When the cameras become unimpaired again, after the user exits the adverse environment, the uncertainty falls back to nominal levels with the acquisition of a subsequent landmark match and the system resumes optimal tracking.

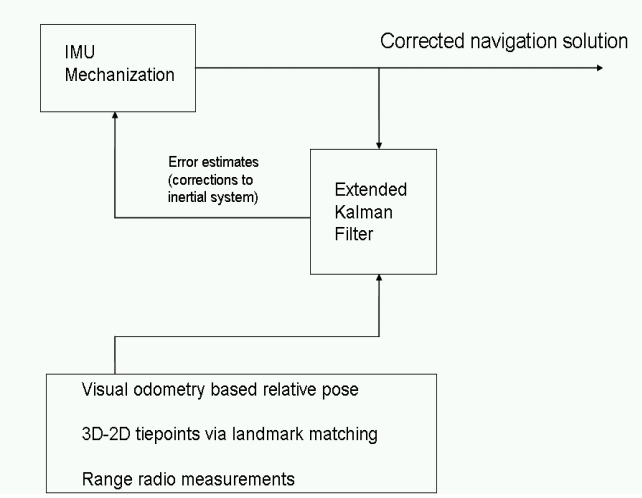

Fig. 2. Error-state Extended Kalman Filter block diagram with local and global external measurements.

II. Extended Kalman Filter Process Model

We denote the ground (global coordinate frame) to camera pose as $\mathbf{P}_{GC} = [\mathbf{R}_{GC} \quad \mathbf{T}_{GC}]$ such that a point $\mathbf{X}_G$ expressed in the ground frame can be transferred to the camera coordinates by $\mathbf{X}_C = \mathbf{R}_{GC}\mathbf{X}_G + \mathbf{T}_{GC}$. Accordingly, $\mathbf{T}_{GC}$ represents the ground origin expressed in the camera

coordinate frame, whereas $\mathbf{T}_{CG} = -\mathbf{R}_{GC}^T\mathbf{T}_{GC}$ is the camera location in the ground coordinate frame.

In this paper, without loss of generality and to keep the notation simple, we will assume that the camera and IMU coordinate systems coincide so that $\mathbf{P}_{GI} = \mathbf{P}_{GC}$. In reality we use an extrinsic calibration procedure to determine the camera to IMU pose $\mathbf{P}_{CI}$, (front left stereo camera is chosen as the master) as developed in [3] and distinguish between $\mathbf{P}_{GI} = \mathbf{P}_{CI}\mathbf{P}_{GC}$ and $\mathbf{P}_{GC}$.

The total (full) states of the filter consist of range radio bias $b_r^{(k)}$ for each node $1 \leq k \leq K$, the camera location $\mathbf{T}_{CG}$, the gyroscope bias vector $\mathbf{b}_g$, velocity vector $\mathbf{v}$ in global coordinate frame, accelerometer bias vector $\mathbf{b}_a$ and ground to camera orientation $\mathbf{q}_{GC}$, expressed in terms of the quaternion representation for rotation, such that $\mathbf{R}_{GC} = (|q_0|^2 - \|\overrightarrow{\mathbf{q}}\|^2)\mathbf{I}_{3\times 3} + 2\overrightarrow{\mathbf{q}}\,\overrightarrow{\mathbf{q}}^T - 2q_0\left[\overrightarrow{\mathbf{q}}\right]_\times$, with $\mathbf{q}_{GC} = [q_0 \;\; \overrightarrow{\mathbf{q}}^T]^T$ and $\left[\overrightarrow{\mathbf{q}}\right]_\times$ denoting the skew-symmetric matrix formed by $\overrightarrow{\mathbf{q}}$. For quaternion algebra, we follow the notation and use the frame rotation perspective as described in [1]. Hence, the total (full) state vector is given by

$$\mathbf{s} = [\mathbf{q}_{GC}^T \quad \mathbf{b}_g^T \quad \mathbf{v}^T \quad \mathbf{b}_a^T \quad \mathbf{T}_{CG}^T \quad b_r^{(1)} \cdots b_r^{(K)}] \; . \quad (1)$$

We use the corresponding system model for the state time evolution

$$\dot{\mathbf{q}}_{GC}(t) = \frac{1}{2}(\mathbf{q}_{GC}(t) \otimes \boldsymbol{\omega}(t)), \quad \dot{\mathbf{b}}_g(t) = \mathbf{n}_{wg}(t)$$
$$\dot{\mathbf{v}}(t) = \mathbf{a}(t), \quad \dot{\mathbf{b}}_a(t) = \mathbf{n}_{wa}(t), \quad \dot{\mathbf{T}}_{CG}(t) = \mathbf{v}(t)$$
$$\dot{b}_r^{(k)}(t) = n_{wr}^{(k)}(t), \quad 1 \leq k \leq K \; ,$$

where $\mathbf{n}_{wg}$, $\mathbf{n}_{wa}$, and $n_{wr}^{(k)}$ for $1 \leq k \leq K$ are modeled as white Gaussian noise, and $\mathbf{a}(t)$ is camera acceleration in global coordinate frame, and $\boldsymbol{\omega}(t)$ is the rotational velocity in camera coordinate frame. Gyroscope and accelerometer measurements of these two vectors are modeled as:

$$\boldsymbol{\omega}_m(t) = \boldsymbol{\omega}(t) + \mathbf{b}_g(t) + \mathbf{n}_g(t) \quad (2)$$
$$\mathbf{a}_m(t) = \mathbf{R}_{GC}(t)(\mathbf{a}(t) - \mathbf{g}) + \mathbf{b}_a(t) + \mathbf{n}_a(t) \quad (3)$$

where $\mathbf{n}_g$ and $\mathbf{n}_a$ are modeled as white Gaussian noise and $\mathbf{g}$ is the gravitational acceleration expressed in the global coordinate frame.

State estimate propagation is obtained by the IMU mechanization equations

$$\dot{\hat{\mathbf{q}}}_{GC}(t) = \frac{1}{2}(\hat{\mathbf{q}}_{GC}(t) \otimes \hat{\boldsymbol{\omega}}(t)) \quad (4)$$
$$\dot{\hat{\mathbf{v}}}(t) = \hat{\mathbf{R}}_{GC}^T(t)\hat{\alpha}(t) + \mathbf{g}, \quad (5)$$
$$\dot{\hat{\mathbf{x}}}(t) = \hat{\mathbf{v}}(t), \qquad \dot{\hat{\mathbf{b}}}_g(t) = 0, \qquad \dot{\hat{\mathbf{b}}}_a(t) = 0 \quad (6)$$

where $\hat{\boldsymbol{\omega}}(t) = \boldsymbol{\omega}_m(t) - \hat{\mathbf{b}}_a(t)$, and $\hat{\alpha}(t) = \mathbf{a}_m(t) - \hat{\mathbf{b}}_a(t)$, together with the radio bias propagation

$$\dot{\hat{b}}_r^{(k)}(t) = 0, \quad 1 \leq k \leq K \; . \quad (7)$$

We solve the above system of equations by fourth-order Runge-Kutta numerical integration method. The Kalman filter error state consists of

$$\delta\mathbf{s} = [\delta\boldsymbol{\Theta}^T \;\; \delta\mathbf{b}_g^T \;\; \delta\mathbf{v}^T \;\; \delta\mathbf{b}_a^T \;\; \delta\mathbf{T}_{CG}^T \;\; \delta b_r^{(1)} \cdots \delta b_r^{(K)}]^T \quad (8)$$

according to the following relation between the total state and its inertial estimate

$$\mathbf{q}_{GC} = \hat{\mathbf{q}}_{GC} \otimes \delta\mathbf{q}_{GC}, \quad \text{with} \;\; \delta\mathbf{q}_{GC} \simeq [1 \quad \frac{\delta\boldsymbol{\Theta}^T}{2}]^T \quad (9)$$
$$\mathbf{b}_g(t) = \hat{\mathbf{b}}_g(t) + \delta\mathbf{b}_g(t), \;\; \mathbf{b}_a(t) = \hat{\mathbf{b}}_a(t) + \delta\mathbf{b}_a(t) \quad (10)$$
$$\mathbf{v}(t) = \hat{\mathbf{v}}(t) + \delta\mathbf{v}(t), \;\; \mathbf{T}_{CG}(t) = \hat{\mathbf{T}}_{CG}(t) + \delta\mathbf{T}_{CG}(t) \quad (11)$$

together with

$$b_r^{(k)}(t) = \hat{b}_r^{(k)}(t) + \delta b_r^{(k)}(t), \quad 1 \leq k \leq K \; , \quad (12)$$

based on which we obtain (after some algebra) the following dynamic process model for the error state:

$$\dot{\delta\mathbf{s}} = \mathbf{F}\delta\mathbf{s} + \mathbf{G}\mathbf{n} \quad (13)$$

where

$$\mathbf{F} = \begin{bmatrix} -[\hat{\boldsymbol{\omega}}]_\times & -\mathbf{I}_3 & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times K} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times K} \\ -\hat{\mathbf{R}}_{GC}^T[\hat{\alpha}]_\times & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & -\hat{\mathbf{R}}_{GC}^T & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times K} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times K} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{I}_3 & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times K} \\ \mathbf{0}_{K\times 3} & \mathbf{0}_{K\times 3} & \mathbf{0}_{K\times 3} & \mathbf{0}_{K\times 3} & \mathbf{0}_{K\times 3} & \mathbf{0}_{K\times K} \end{bmatrix}$$

and

$$\mathbf{G} = \begin{bmatrix} -\mathbf{I}_3 & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times K} \\ \mathbf{0}_{3\times 3} & \mathbf{I}_3 & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times K} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & -\hat{\mathbf{R}}_{GC}^T & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times K} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{I}_3 & \mathbf{0}_{3\times K} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times K} \\ \mathbf{0}_{K\times 3} & \mathbf{0}_{K\times 3} & \mathbf{0}_{K\times 3} & \mathbf{0}_{K\times 3} & \mathbf{I}_K \end{bmatrix}$$

and

$$\mathbf{n} = [\mathbf{n}_g^T \quad \mathbf{n}_{wg}^T \quad \mathbf{n}_a^T \quad \mathbf{n}_{wa}^T \quad n_{wr}^{(1)} \cdots n_{wr}^{(K)}]^T \; .$$

III. Visual Odometry and Landmark Matching Measurement Model

To incorporate visual odometry poses that are relative in nature, we apply the same stochastic cloning approach developed in [5] for our measurement model. In particular, we denote $\mathbf{P}_{1,2}$ as the visual odometry pose estimate between two time instances 1 and 2, and let the corresponding pose components of the state be denoted by $\mathbf{P}_{G,1}$ and $\mathbf{P}_{G,2}$. Then defining $\mathbf{T}_{2,1} = \mathbf{R}_{G,1}(\mathbf{T}_{2,G} - \mathbf{T}_{1,G})$, and $\mathbf{q}_{1,2} = \mathbf{q}_{G,1}^{-1}\,\mathbf{q}_{G,2}$, and after lengthy algebra as similar to [5], we obtain the following measurement equations

$$\delta\mathbf{z}_T = \left[\hat{\mathbf{R}}_{G,1}(\hat{\mathbf{T}}_{2,G} - \hat{\mathbf{T}}_{1,G})\right]_\times \delta\boldsymbol{\Theta}_{G,1} + \hat{\mathbf{R}}_{G,1}\delta\mathbf{T}_{2,G} \quad (14)$$
$$- \hat{\mathbf{R}}_{G,1}\delta\mathbf{T}_{1,G} + \boldsymbol{\nu}_T \quad (15)$$

and

$$\delta\mathbf{z}_q = 1/2\,\hat{\mathbf{R}}_{1,2}^T\delta\boldsymbol{\Theta}_{G,2} - 1/2\,\delta\boldsymbol{\Theta}_{G,1} + \boldsymbol{\nu}_q \quad (16)$$

where $\boldsymbol{\nu}_T$ and $\boldsymbol{\nu}_q$ are the Gaussian noise in translation and rotation associated with the visual odometry pose solution. These measurements are a function of the propagated error-state $\delta\mathbf{s}_2$ and the cloned error-state $\delta\mathbf{s}_1$ from previous time instance, which require modifications to the Kalman filter update equations (cf. [5]).

As for landmark matching, given a query image, landmark matching returns the found landmark shot from the database establishing the 2D to 3D point correspondences between the query image features and the 3D local point cloud, as well as the camera pose $\mathbf{P}_{GL}$ belonging to that shot. First, every 3D local landmark point $\mathbf{X}$ is transferred to the global coordinate system via

$$\mathbf{Y} = \mathbf{R}_{LG}\mathbf{X} + \mathbf{T}_{LG} \tag{17}$$

which can be written under small error assumption as

$$\hat{\mathbf{Y}} + \delta\mathbf{Y} \simeq (\mathbf{I} - [\boldsymbol{\rho}]_\times)\hat{\mathbf{R}}_{LG}(\hat{\mathbf{X}} + \delta\mathbf{X}) + \hat{\mathbf{T}}_{LG} + \delta\mathbf{T}_{LG}$$

where $\boldsymbol{\rho}$ is a small rotation vector. Neglecting second order terms results in the following linearization

$$\delta\mathbf{Y} \simeq \hat{\mathbf{R}}_{LG}\delta\mathbf{X} + \left[\hat{\mathbf{R}}_{LG}\hat{\mathbf{X}}\right]_\times \boldsymbol{\rho} + \delta\mathbf{T}_{LG} \tag{18}$$

and letting $\tilde{\mathbf{X}} = \hat{\mathbf{R}}_{LG}\hat{\mathbf{X}}$, the local 3D point covariance, $\boldsymbol{\Sigma}_Y$, can be represented in the global coordinate frame in terms of the local reconstruction uncertainty, $\boldsymbol{\Sigma}_X$ and landmark pose uncertainty in rotation and translation, $\boldsymbol{\Sigma}_{\mathbf{R}_{LG}}$ and $\boldsymbol{\Sigma}_{\mathbf{T}_{LG}}$, as

$$\boldsymbol{\Sigma}_Y \simeq \hat{\mathbf{R}}_{LG}\boldsymbol{\Sigma}_X\hat{\mathbf{R}}_{LG}^T + [\tilde{\mathbf{X}}]_\times\boldsymbol{\Sigma}_{\mathbf{R}_{LG}}[\tilde{\mathbf{X}}]_\times^T + \boldsymbol{\Sigma}_{\mathbf{T}_{LG}}$$

After this transformation, the projective camera measurement model is employed such that for each 3D point $\mathbf{Y}$ obtained above and expressed in the current camera coordinate system as $\mathbf{Z} = [Z_1 \;\; Z_2 \;\; Z_3]^T$, the projection onto the normalized image plane is given by

$$\mathbf{z} = f(\mathbf{Z}) + \boldsymbol{\nu} \;\; \text{with} \;\; f(\mathbf{Z}) = [Z_1/Z_3 \;\; Z_2/Z_3]^T \tag{19}$$

where $\boldsymbol{\nu}$ is the feature measurement noise with covariance $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}$ and

$$\mathbf{Z} = \mathbf{R}_{GC}\mathbf{Y} + \mathbf{T}_{GC} = \mathbf{R}_{GC}(\mathbf{Y} - \mathbf{T}_{CG}) \; . \tag{20}$$

Under small error assumption

$$\hat{\mathbf{Z}} + \delta\mathbf{Z} \simeq (\mathbf{I} - [\delta\boldsymbol{\Theta}]_\times)\hat{\mathbf{R}}_{GC}(\hat{\mathbf{Y}} + \delta\mathbf{Y} - \hat{\mathbf{T}}_{CG} - \delta\mathbf{T}_{CG}) \; .$$

Hence,

$$\delta\mathbf{Z} \simeq \left[\hat{\mathbf{R}}_{GC}(\hat{\mathbf{Y}} - \hat{\mathbf{T}}_{CG})\right]_\times \delta\boldsymbol{\Theta} + \hat{\mathbf{R}}_{GC}(\delta\mathbf{Y} - \delta\mathbf{T}_{CG}) + \boldsymbol{\nu} \; .$$

Accordingly, the measurement equation in the error-states is given by

$$\delta\mathbf{z}_L \simeq \mathbf{H}_L\delta\mathbf{s} + \boldsymbol{\eta} \tag{21}$$

where the measurement Jacobian

$$\mathbf{H}_L = [\mathbf{J}_f\mathbf{J}_{\boldsymbol{\Theta}} \quad \mathbf{0}_{3\times3} \quad \mathbf{0}_{3\times3} \quad \mathbf{0}_{3\times3} \quad \mathbf{J}_f\mathbf{J}_{\delta\mathbf{T}_{CG}}] \tag{22}$$

with

$$\mathbf{J}_f = \begin{bmatrix} 1/\hat{Z}_3 & 0 & -\hat{Z}_1/\hat{Z}_3^2 \\ 0 & 1/\hat{Z}_3 & -\hat{Z}_2/\hat{Z}_3^2 \end{bmatrix} \tag{23}$$

$$\mathbf{J}_{\boldsymbol{\Theta}} = \left[\hat{\mathbf{R}}_{GC}(\hat{\mathbf{Y}} - \hat{\mathbf{T}}_{CG})\right]_\times, \;\; \text{and} \;\; \mathbf{J}_{\delta\mathbf{T}_{CG}} = -\hat{\mathbf{R}}_{GC}$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta}} = \hat{\mathbf{R}}_{GC}\boldsymbol{\Sigma}_Y\hat{\mathbf{R}}_{GC}^T + \boldsymbol{\Sigma}_{\boldsymbol{\nu}} \tag{24}$$

The above is applied to all the point correspondences returned as a result of landmark matching, and all the matrices and vectors are stacked to form the final measurement model equation.

IV. RF Range Measurement Model

Each radio node provides a measurement of its range to every other node in the system, which we model as

$$z_r^{(k)} = \|\mathbf{T}_{CG} - \mathbf{T}_{RG}^{(k)}\| + b_r^{(k)} + \nu_r^{(k)}, \quad 1 \le k \le K \tag{25}$$

where $\nu_r^{(k)}$ is white Gaussian measurement noise and we denote by $\mathbf{T}_{RG}^{(k)}$, the location in global coordinates of that particular node that is being ranged to, whose location is known by the radio that is doing the ranging. (Note that, coordinates of the static nodes are stored in each unit and remain fixed, whereas the location of mobile nodes are given by $\mathbf{T}_{CG}^{(k)}$ and are continuously broadcast at the frame rate over the wireless network.) Using the small error assumption, the above can be written as

$$\hat{z}_r^{(k)} + \delta z_r^{(k)} \simeq \|\hat{\mathbf{T}}_{CG} - \mathbf{T}_{RG}^{(k)}\| + \mathbf{J}_r\delta\mathbf{T}_{CG} + \hat{b}_r^{(k)} + \delta b^{(k)} + \nu_r^{(k)} \tag{26}$$

where

$$\mathbf{J}_r = \frac{\hat{\mathbf{T}}_{CG}^T - {\mathbf{T}_{RG}^{(k)}}^T}{\|\hat{\mathbf{T}}_{CG} - \mathbf{T}_{RG}^{(k)}\|}$$

so that

$$\delta z_r^{(k)} \simeq \mathbf{J}_r\delta\mathbf{T}_{CG} + \delta b^{(k)} + \nu_r^{(k)} \tag{27}$$

from which we have

$$\delta z_r^{(k)} = \mathbf{H}_r\delta\mathbf{s} + \nu_r^{(k)} \tag{28}$$

with

$$\mathbf{H}_r = [\mathbf{0}_{1\times3} \quad \mathbf{0}_{1\times3} \quad \mathbf{0}_{1\times3} \quad \mathbf{0}_{1\times3} \quad \mathbf{J}_r \quad \mathbf{e}_{1\times K}^{(k)}] \; , \tag{29}$$

where $\mathbf{e}_{1\times K}^{(k)}$ is a $1 \times K$ vector of zeros whose k'th element is a 1.

V. Experimental Results

We have done extensive tests of our current system around our campus and here we present some results from these experiments. First, we demonstrate results from a single user experiment where we have not created any challenging conditions to break visual tracking and no range radios are used. Fig. 3 shows the automatically generated real-time camera trajectory for this run corresponding to an 810 meter course within Sarnoff campus completed by a user wearing our helmet and backpack system and walking indoors and outdoors in several loops. The entire area shown in the map is within the pre-built landmark database capture range which is loaded in the beginning before the exercise takes place and landmark matches occur whenever a query image is within close proximity to a stored landmark shot in the database. Fig. 4 shows several screen shots corresponding to locations towards the beginning, middle and end of this exercise obtained from our visualization tool which we use to verify the accuracy of the camera pose outputs. It is observed that these views are in very good agreement which indicate how precisely the camera is tracked throughout the entire duration of the course.

Next, results from a more complex three-user experiment are shown. In this case, we used five range radios as depicted

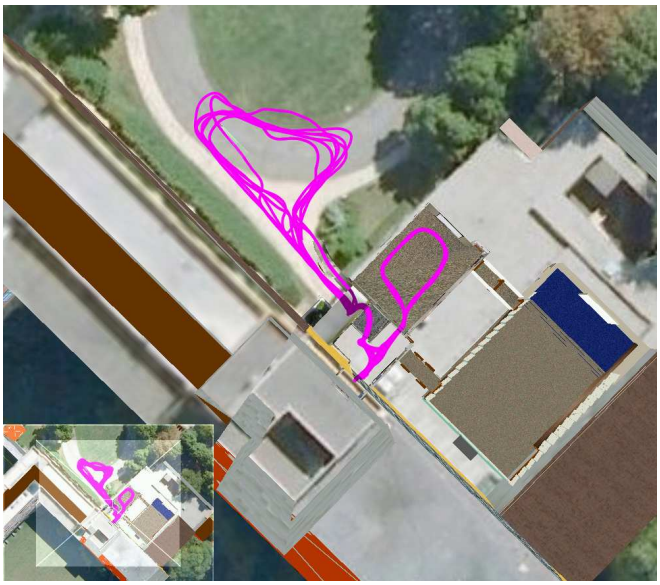

Fig. 3. Real-time computed camera trajectory corresponding to a 810 meter long course completed in 16.4 minutes during an online exercise.

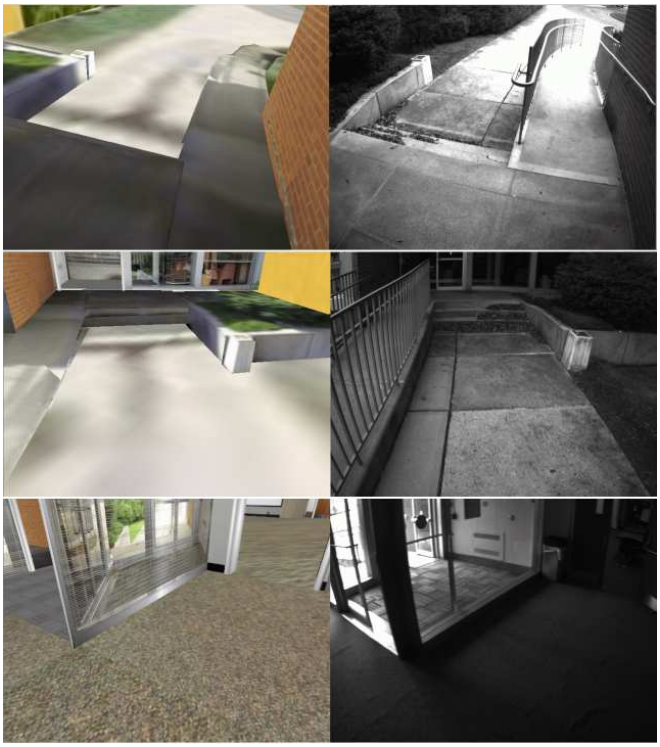

Fig. 4. The views rendered from the model using the real-time camera pose estimates by our system for various locations throughout the exercise, together with the real scene views captured by the camera.

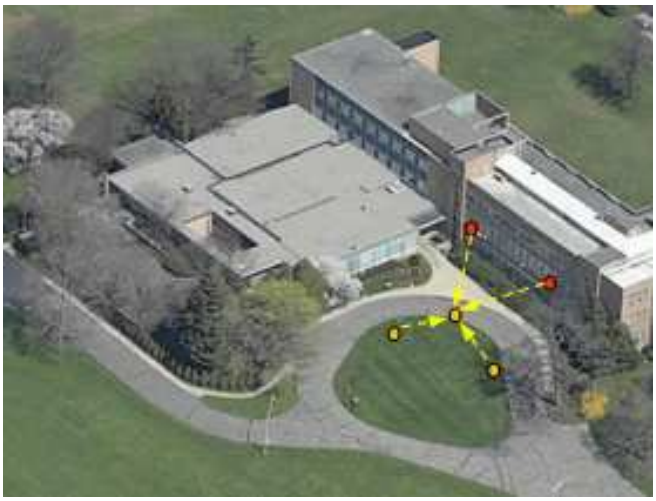

Fig. 5. Static anchor node locations in the map are shown in red corresponding to radios placed on window sills on the second floor of the building. Mobile nodes attached to backpack systems worn by three users during live exercise are shown in yellow.

Fig. 6. Rendered views from the model together with the real scene views captured by the camera before the user enters the smoke region and right after he exits.

in Fig. 5. (Supplemental video taken during the exercise is provided.) In this exercise, we have used a smoke machine to block the visual cues and one user continually walks in and out of the smoke covered region (cf. Fig. 6). In the video, the trajectory corresponding to this user can be seen in our "mapview" visualization tool as it is displayed live during the exercise. In Fig. 7, we show the trajectories of the three users participating in the exercise. The magenta colored path belongs to the user going through the smoke while the blue and green paths belong to those users that stay out of the smoke and act as mobile beacons.

Fig. 8 shows the range measurements, in blue, from one mobile node to all the other four nodes, and in red, the corresponding Kalman filter estimates are shown. One can see that there is close agreement between the filter estimates

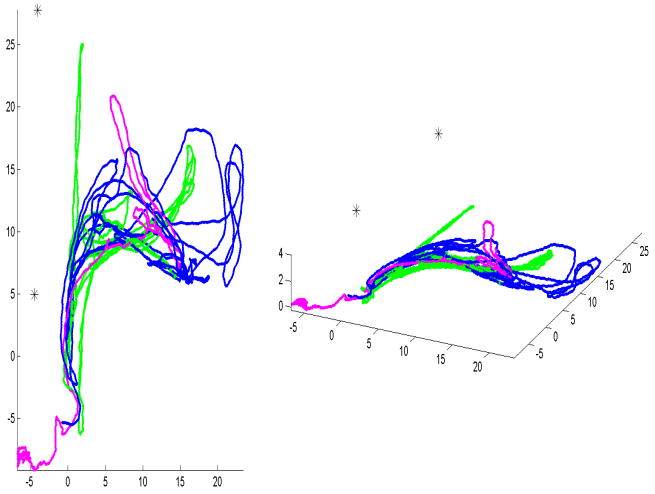

Fig. 7. Top and oblique views showing the trajectories of three users in a live exercise. All axes are in meters. The path in magenta corresponds to the user who continuously enters in the smoke covered region. The green and blue paths belong to the other two users who act as mobile beacons. The location of the static anchor radio nodes are shown in asterisks.

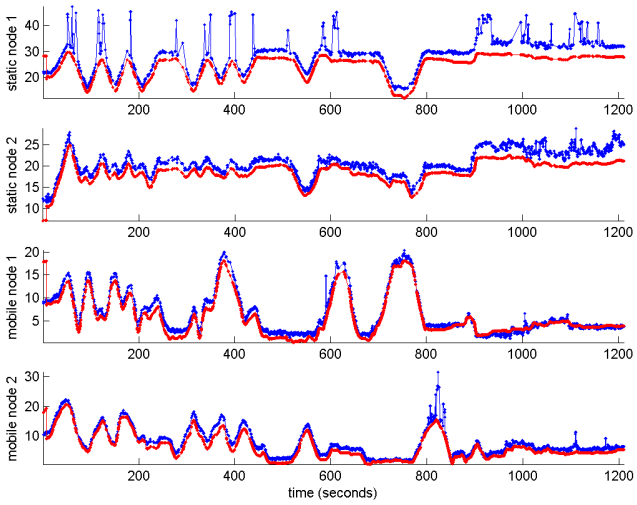

Fig. 8. Ranges to each node in meters as a function of time. Top two plots show the ranges of the mobile node of interest to the static anchor nodes, whereas the bottom two plots depict the ranges to the other two mobile nodes. In blue are the raw radio measurements and in red are the estimates output by the Kalman filter.

and radio measurements. The gaps between each of these are compensated by the radio bias components which are also tracked as part of the filter state and shown in Fig. 9. Upon close inspection, one can notice that for a brief period in the beginning, the ranges in red are very far from the radio measurements. This is because during this period, the user has not acquired a landmark match which is needed to place the filter state in the global coordinate system. Starting with the first match, the pose component of the state is reset to the pose returned by the landmark match module. The radio measurements are not used inside the filter until after this has occurred. This takes place in all the mobile nodes and each mobile node does not transmit its location prior to having

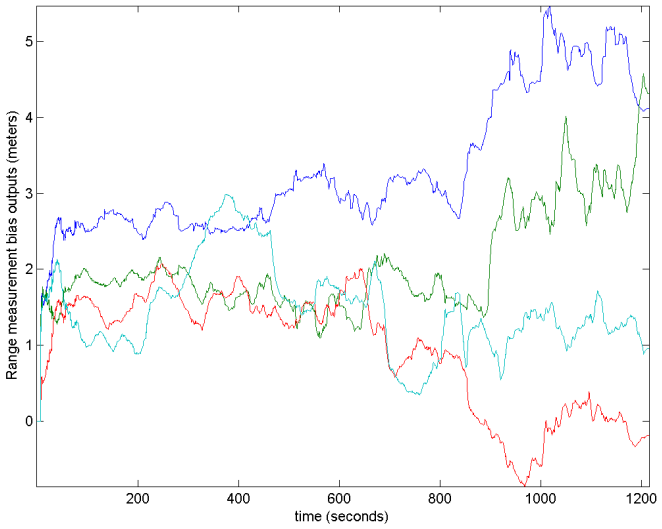

Fig. 9. Radio range measurement biases to each of the four nodes, as tracked by the Kalman filter.

acquired its first landmark match.

VI. Conclusion

We presented a unified Kalman filter framework using local and global sensor data fusion for vision aided navigation related to augmented reality and training applications and showed results to illustrate the accuracy and robustness of our system over long duration and distance. Using a pre-built landmark database of the entire exercise area and employing range radio measurements from both static and mobile nodes eliminate the problem of long term drift inherent in any inertial based navigation platform and provide ubiquitous and precise tracking both indoors and outdoors and under challenging conditions for a vision-based localization approach.

References

[1] J. B. Kuipers. *Quaternions and Rotation Sequences*. Princeton University Press, 1998.

[2] F. Lu and E. Milios. Globally consistent range scan alignment for environment mapping. *Autonomous Robots*, 4, 1997.

[3] F. M. Mirzaei and S. I. Roumeliotis. A kalman filter-based algorithm for imu-camera calibration: Observability analysis and performance evaluation. *IEEE Transactions on Robotics*, 24(5), 2008.

[4] T. Oskiper, Z. Zhu, S. Samarasekera, and R. Kumar. Visual odometry system using multiple stereo cameras and inertial measurement unit. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[5] S. I. Roumeliotis, A. E. Johnson, and J. F. Montgomery. Augmenting inertial navigation with image-based motion estimation. In *IEEE International Conference on Robotics and Automation*, 2002.

[6] S. I. Roumeliotis, G. S. Sukhatme, and G. Bekey. Circumventing dynamic modeling: Evaluation of the error-state kalman filter applied to mobile robot localization. In *IEEE International Conference on Robotics and Automation*, 1999.

[7] S. Scheding, E. M. Nebot, S. M., and D.-W. H. F. Experiments in autonomous underground guidance. In *IEEE International Conference on Robotics and Automation*, 1997.

[8] S. Thrun, D. Fox, W. Burgard, and F. Dellaert. Robust monte carlo localization for mobile robots. *J. Artificial Intelligence*, 101, 2001.

[9] Z. Zhu, T. Oskiper, S. Samarasekera, R. Kumar, and H. S. Sawhney. Real-time global localization with a pre-built visual landmark database. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.