2013 IEEE International Conference on Robotics and Automation (ICRA)
Karlsruhe, Germany, May 6-10, 2013

Adaptive-Frame-Rate Monocular Vision and IMU Fusion for Robust Indoor Positioning

Ya Tian, *Student Member, IEEE*, Jie Zhang, and Jindong Tan, *Member, IEEE*

***Abstract*— Robust navigation for mobile robots requires an accurate method for tracking the robot position in the environment. This paper presents a simple and novel visual-inertial integration system suitable for unstructured and unprepared indoor environments, where MARG (Magnet, Angular Rate and Gravity) sensors and a monocular camera are used. The pre-estimated orientation from MARG sensors, is used to estimate the translation based on the data from the visual and inertial sensors. This has a significant effect on the performance of the fusion sensing strategy and makes the fusion procedure much easier, because the gravitational acceleration can be correctly removed from the accelerometer measurements before the fusion procedure, where a linear Kalman Filter is selected as the fusion estimator. the use of pre-estimated orientation can help to eliminate erroneous point matches based on the properties of the pure camera translation and thus the computational requirements can be significantly reduced compared to the RANSAC (RANdom SAmple Consensus) algorithm. In addition, an adaptive-frame-rate single camera is selected to not only avoid motion blur based on the angular velocity and acceleration after compensation but also to make an effect called visual zero-velocity update for the static motion. Thus, it can recover a more accurate baseline and meanwhile reduce the computational requirements. In particular, an absolute scale factor, which is usually lost in monocular camera tracking, can be obtained by introducing it into the estimator. Simulation and experimental results are presented for different environments with different types of movement and the results from a Pioneer robot are used to demonstrate the accuracy of the proposed method.**

I. INTRODUCTION

Robust positioning plays an important role in navigation systems. It is known that inertial sensors are popularly used in navigation systems, called Inertial Navigation System (INS) [1] [2], because they can sample data in high-frequency rate and give the precise results in short term without constrained operation. Moreover, they can be light weight, low cost, and small size and adopt wireless communication technologies, which make the attachment of them on human body much easier. That is why we call them ”wearable” inertial sensors. However, directly double integrating the linear acceleration directly from inertial sensors will lead to long-term drift accumulated in position estimates. It is also a major issue that limits the accuracy and performance of inertial navigation. In addition, the effect of the gravity, called the gravitational acceleration, should be considered and correctly removed from the accemlerometer measurements. Therefore, some form of aiding is normally required to maintain the integrity of the pose information. In order to solve this problem, many techniques based on the Global Positioning System (GPS) (mainly used in outdoor environments) have been proposed by researchers in combination with inertial systems to overcome the long-term error growth [3]–[5]. However, the major problem of the GPS signal is intermittent loss in navigation based on GPS-aided inertial systems, especially for indoor navigation. Hence, an alternative technology is needed to ensure smooth and reliable inertial navigation during GPS signal outages.

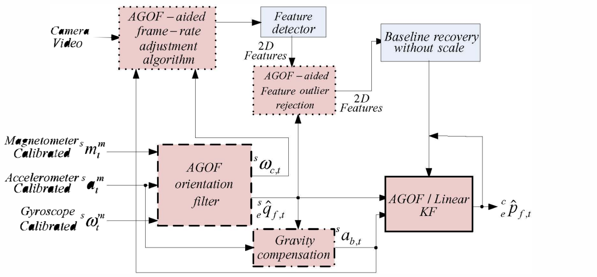

Camera Video
AGOF – aided frame – rate adjustment algorithm
Feature detector
2D Features
AGOF – aided Feature outlier rejection
2D Features
Baseline recovery without scale
Magnetometer Calibrated $^{s}m_t^m$
Accelerometer Calibrated $^{s}a_t^m$
Gyroscope Calibrated $^{s}\omega_t^m$
AGOF orientation filter
$^{s}\omega_{c,t}$
$^{s}_{e}\hat{q}_{f,t}$
Gravity compensation
$^{s}a_{b,t}$
AGOF / Linear KF
$^{c}_{e}\hat{p}_{f,t}$

Fig. 1. The main framework of the proposed method for position estimation.

The integration between visual and inertial sensors is motivated by what happens with the vestibular (inner ear) and vision (eyes) system in humans and animals [6], and the basic principle for integrating vision and inertial sensors together is their complementary properties. As we aforementioned, inertial sensors offer good signals with high rate during fast motions but are sensitive to accumulated drift due to the double integration during the estimation of position. On the contrary, visual sensors provide precise ego-motion estimation with low rate in long term, but suffer from blurred features under fast and unpredicted motions [7]. The aim of visual-inertial sensor integration is to overcome some fundamental limitations of vision-only tracking and IMU-only (Inertial Measurement Unit) tracking using their complementary properties.

Even though the motivation of fusing visual and inertial sensors is their complementary properties, the following challenges need to be solved for more robust fusion results. First, the major disadvantage of inertial sensors is the accumulated drift, which should be reduced by not only using visual sensors but also using the properties of inertial sensors themselves, such as the gravity from the accelerometer and the earth magnetic north from the magnetometer. Second, obtaining accurate readings from the accelerometer is the key point for the accurate position estimation, which needs to be calculated by double integration of the acceleration. Third, visual feature correspondences are easily lost due to motion blur caused by fast camera motion, especially

Y. Tian is with the School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan, Shandong, 250101, China. e-mail: {tya}@mtu.edu. Jie Zhang is with Beijing University of Chemical Technology and is a visiting scholar at the University of Tennessee, Knoxville. J. Tan is with the Department of Mechanical, Aerospace and Biomedical Engineering, The University of Tennessee, Knoxville, TN 37996, USA. e-mail: {tan}@utk.edu.

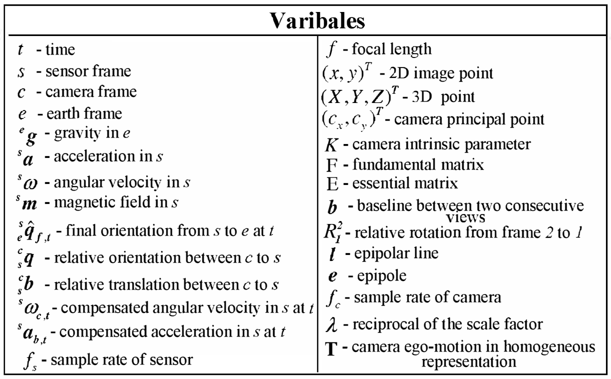
978-1-4673-5643-5/13/$31.00 ©2013 IEEE

fast rotation. And lastly, the fusion methods usually adopt iterative algorithms to deal with non-linear models, and hence convergence is not always assured [7]–[9]. In addition, the tracking environment is usually predefined with specific objects or landmarks [7], [9].

In order to overcome these challenges, 5 strategies are proposed as the major parts of the main framework, shown as in Figure 1: (a) AGOF orientation filter; (b) gravity compensation; (c) AGOF-aided frame-rate adjustment algorithm; (d) AGOF-aided feature outlier rejection; and (e) AGOF/Linear Kalman Filter. The last three parts are also the main contributions of this paper.

II. Problem Formulation

This section formulates the orientation and position estimation problem using inertial and visual measurements. In order to better describe the problem, a quaternion representation of orientation for the sensor arrays is used throughout this paper (see references in [10]–[12]). For brevity and clarity, the definitions of mathematical variables and symbols are firstly described and shown in Figure 2.



Varibales	
t - time	f - focal length
s - sensor frame	$(x,y)^T$ - 2D image point
c - camera frame	$(X,Y,Z)^T$ - 3D point
e - earth frame	$(c_x,c_y)^T$ - camera principal point
${}^e\boldsymbol{g}$ - gravity in e	K - camera intrinsic parameter
${}^s\boldsymbol{a}$ - acceleration in s	F - fundamental matrix
${}^s\boldsymbol{\omega}$ - angular velocity in s	E - essential matrix
${}^s\boldsymbol{m}$ - magnetic field in s	$\boldsymbol{b}$ - baseline between two consecutive views
${}^s_e\hat{\boldsymbol{q}}_{f,t}$ - final orientation from s to e at t	R_1^2 - relative rotation from frame 2 to 1
${}^c_s\boldsymbol{q}$ - relative orientation between c to s	$\boldsymbol{l}$ - epipolar line
${}^c_s\boldsymbol{b}$ - relative translation between c to s	$\boldsymbol{e}$ - epipole
${}^s\boldsymbol{\omega}_{c,t}$ - compensated angular velocity in s at t	f_c - sample rate of camera
${}^s\boldsymbol{a}_{b,t}$ - compensated acceleration in s at t	λ - reciprocal of the scale factor
f_s - sample rate of sensor	$\mathbf{T}$ - camera ego-motion in homogeneous representation

Fig. 2. Symbols: the left side shows symbols from inertial sensors and the right side shows symbols from camera

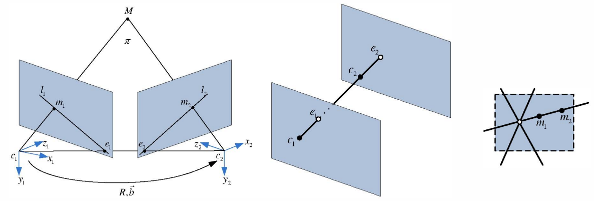
A. MARG Sensors

Based on the limitation of the MARG sensor, and the fact that the proposed algorithm is implemented on a newly developed low-cost chip [13] shown in Figure 5(a), the measurement models for the gyroscope measurements ${}^s\omega_t^m$, accelerometer measurements ${}^sa_t^m$, and magnetometer measurements ${}^sm_t^m$ can be found in [10]. These measurements are also inputs of our AGOF filter as shown in Figure 1. Calibration procedure for the MARG sensors should be taken before each practical use, unless the characteristics of the sensors themselves change little. In this paper, the *null point* and *scale factor* of each axis of the MARG sensors are determined using the method described in [14]. After sensor modeling and calibration, the final quaternion-based orientation ${}^s_e\hat{q}_{f,t}$, compensated acceleration ${}^sa_{b,t}$, and compensated angular velocity ${}^s\omega_{c,t}$ can be obtained for position estimation.

B. Visual Sensors

The basic pinhole camera model [15] is selected in this paper to infer the mapping between a 2D point in pixel coordinates and a 3D point in camera coordinates, which can be easily expressed as follows using the camera's intrinsic calibration matrix K [15].

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = s * \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = s * K * \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$



(a) Epipolar Geometry (b) Pure Translation

Fig. 3. (a) Epipolar geometry between two consecutive frames. (b) The left image shows the pure translation from the camera center c_1 to c_2. e_2, which expresses the epipole in current frame, is the vanishing point of the direction of translation. The right image shows the property of pure translational movement with the parallel image planes and matched feature points radiating from the epipole.

where f_x and f_y are the so-called focal distances, which are derived from the focal length f multiplied by k_x and k_y respectively (k_x and k_y denote the number of pixels per unit of length ([pixel/length]) in x and y direction [15]).

After camera modeling and calibration, the baseline b between two consecutive images can be obtained for position estimation.

C. Hand-Eye Calibration

The visual and inertial sensors are rigidly attached, so the rotation and translation between them are constant. Usually, the hand-eye calibration is the task of computing the relative 3D position and orientation between the camera and the IMU in an eye-on-hand configuration, where the camera is rigidly connected to the inertial sensors. The method proposed in [16] is used to determine the relative rotation ${}^c_sq$ and translation ${}^c_sb$ between these two coordinate frames.

III. Adaptive Frame-rate Monocular Camera Tracking

The goal of this work is to recover more accurate relative translation between two consecutive frames called *baseline b*, and meanwhile reduce the computational requirements, when severe rotation or translation happens. Therefore, an adaptive frame-rate algorithm is proposed to achieve this goal.

A. AGOF-aided Feature Outlier Rejection

It is understood that the problem of inevitable incorrect correspondences would exist in the group of point correspondences, which would lead to the wrong ego-motion estimation. To reduce this ambiguity and fully use the robust orientation from our AGOF filter, an outlier rejection algorithm based on the pure camera translation after the rotation compensation is proposed to prune the unsuitable correspondences.

1) Epipolar Geometry and Epipoles: As shown in the Figure 3(a), the 3D point M has images m_1 and m_2 at two consecutive time instants. It is clear that m_1 and m_2 are in correspondence if and only if the three vectors c_1m_1, c_2m_2, and c_1c_2 are coplanar on a common epipolar plane π. e_1 and e_2 are the epipoles and l_1 and l_2 are epipolar lines corresponding to the image point m_2 and m_1 respectively.

$$m_2^T \mathrm{F} m_1 = 0 \tag{2}$$

where F is the fundamental matrix, which can be computed using the eight point algorithm [15].

Figure 3(b) shows the property of the pure translational movement of the origin of the camera. Static points move along lines in the image plane. In fact, all lines can intersect at the same point, called "epipole" [15], which can be possibly estimated by tracking feature points on the image plane and computing their lines of movement and intersection.

ixed' camera ($K_1 = K_2 = K$ and $R = \mathrm{I}$), F can be represented as follows:

$$\begin{aligned} \mathrm{F} &= K_2^{-T}[b]_\times R K_1^{-1} \\ &= K^{-T}[b]_\times K^{-1} = [Kb]_\times = [e_2]_\times = [e_1]_\times \end{aligned} \tag{3}$$

where K is the intrinsic parameter of the camera, b is the camera baseline from c_1 to c_2, and R is the camera rotation from c_1 to c_2.

However, the camera generally does not experience only pure translation. Then both of the rotational and translational components are contained in the movement of a feature point on the image plane. Therefore, based on the pre-estimated orientation from our AGOF filter, the rotational information can be easily provided and used for parallel image planes. In this paper, the property of the pure camera translation is used to eliminate incorrect correspondences based on (2) and (3), as shown in (4).

$$\begin{aligned} (R_{c1}^{c2} m_2)^T \mathrm{F} m_1 &= 0 \\ (R_{c1}^{c2} m_2)^T [e]_\times m_1 &= 0 \end{aligned} \tag{4}$$

where R_{c1}^{c2} is the relative rotation between two consecutive frames, obtained from the previous proposed orientation algorithm in [10].

In practice, the point correspondences from the SIFT algorithm contain noise so that (4) can not be satisfied with a small error. Therefore, the smallest reprojection error of SIFT correspondences should be defined. Here, the corresponding epipolar lines $l_1 = ([e]_\times)^T (R_{c1}^{c2} m_2)$ and $l_2 = [e]_\times m_1$ are used to evaluate the distance d shown in (5). Those corresponding features that satisfy (5) will be perfectly consistent with epipolar geometry. Otherwise, they will be considered as *outliers*.

$$d^2 = \frac{((R_{c1}^{c2} m_2)^T [e]_\times m_1)^2}{l_1(1,:)^2 + l_1(2,:)^2 + l_2(1,:)^2 + l_2(2,:)^2} < \varepsilon \tag{5}$$

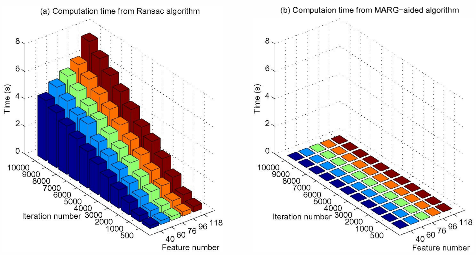


Fig. 4. (a) Computation time from RANSAC algorithm based on the number of iteration and feature. (b) Computation time from MARG-aided outlier rejection algorithm based on the number of iteration and feature.

Figure 4 shows the computation time between the RANSAC algorithm and the proposed AGOF-aided outlier rejection algorithm based on the number of iterations and feature points. As we know, the time computed from the RANSAC heavily depends on the number of iterations and feature points, so a disadvantage of the RANSAC is that there is no upper bound on the time it takes to get the optimal parameters. Figure 4 shows the computation time between the RANSAC algorithm and the proposed AGOF-aided outlier rejection algorithm. Although the proposed algorithm only performs one iteration for each feature point, the figure is still drawn based on the number of iterations and feature points for better comparison with the RANSAC.

B. AGOF-aided Camera Frame-rate Adjustment

In order to obtain an accurate baseline b between two consecutive views based on the epipolar geometry, the movement of the camera between two consecutive views should be significantly large. Therefore, a low frame-rate is selected as the primary frame-rate of our monocular camera for this purpose. However, when the camera goes through fast motion, the low frame-rate of the camera could make images blurred and thus lead to incorrect baseline recovery, so the frame-rate should be increased for avoiding blurred images.

Therefore, an algorithm is developed to sort the camera motion into three types(fast, slow and static motion) based on ${}^s\omega_{c,t}$, ${}^s a_{b,t}$ and 2D point correspondences between two consecutive frames. If the camera moves fast, then increase the frame-rate of the camera; otherwise, keep the pre-defined frame-rate of the camera. Moreover, the static motion can be determined based on 2D point correspondences between two consecutive frames under slow movement, so the vision system can make an effect, called ZUPT, which is commonly used in pedestrian indoor navigation [17]. Algorithm 1 shows the procedure of the proposed AGOF-aided camera frame-rate adjustment structure.

C. Baseline Recovery for Camera Ego-motion without Metric Scale

In [15], the camera matrices may be retrieved from the essential matrix E based on $\mathrm{E} = [b]_\times R$, where $[b]_\times$ is a corresponding 3×3 skew-symmetric matrix of the baseline b and R is the rotation. Therefore, E should be firstly calculated by using the eight-point algorithm [15].

Based on the method proposed in [18], both of the rotation R and the baseline b shown in Figure 3(a) can be recovered by using the (6) and (7) from E.

$$\begin{aligned} b\ b^T &= \tfrac{1}{2} Trace(\mathrm{E}\mathrm{E}^T)\mathrm{I} - \mathrm{E}\mathrm{E}^T \\ (b \cdot b)R &= \mathrm{E}^{*T} - b \times \mathrm{E} \end{aligned} \tag{6}$$

where I is the 3×3 identity matrix and E^* is the matrix of cofactors of E. b is obtained by selecting the largest row of

Algorithm 1 AGOF-aided camera frame-rate adjustment algorithm

```
 0: Set $f_s$, $f_{c_{old}}$ and a predefined M-frame buffer
 1: $f_c = f_{c_{old}}$
 2: Sampling for MARG and camera based on $f_s$ and $f_c$
 3: Data processing based on algorithm for MARG
 4: If $\|{}^s\omega_{c,t}\| \le \varepsilon_\omega$ and $\|{}^s a_{b,t}\| \le \varepsilon_a$ Then
 5:     Capture current frame based on $f_c = f_{c_{old}}$ and load
        the previous frame from M-frame buffer
 6:     2D feature detector using SIFT
 7:     If $\|im_1 \cdot x_1 - im_2 \cdot x_2\| \le \varepsilon_c$ Then
 8:         $b = [0\ 0\ 0]$ and $v = [0\ 0\ 0]$
 9:         Return to step 2
10:     Else
11:         Recover $b$ based on 2D point correspondences
12:         Return to step 2
13:     End If
14: Else If $\|{}^s\omega_{c,t}\| > \varepsilon_\omega$
15:         $f_c = f_{c_{new}} = floor(\frac{\|{}^s\omega_{c,t}\|}{\varepsilon_\omega})$
16:         Capture current frame based on $f_c = f_{c_{new}}$ and
            saved into predefined M-frame buffer
17:         Recover $b$ based on 2D point correspondences
18:         Return to step 2
19:     Else
20:         $f_c = f_{c_{new}} = floor(\frac{\|{}^s a_{b,t}\|}{\varepsilon_a})$
21:         Capture current frame based on $f_c = f_{c_{new}}$ and
            saved into predefined M-frame buffer
22:         Recover $b$ based on 2D point correspondences
23:         Return to step 2
24:     End If
25: End If
```

the matrix $\mathrm{B} = b\ b^T$ and dividing by the square root of the diagonal. Here, b has two solutions (b_1 and b_2) and can be calculated as follows:

$$b \quad = \quad \pm\mathrm{B}(i,:)^T/\sqrt{\mathrm{B}(i,i)} \qquad (7)$$

where $i = 1,2,3$ and $\mathrm{B}(i,i)$ is the largest element of the diagonal of the matrix B. Choosing the right solution can be performed by checking if a reconstructed point has positive depths in front of two views of camera [15].

As we mentioned, b represents relative translation between two consecutive frames, so the absolute ego-motion of the camera at time $t = k$ needs to be calculated based on (8) assuming the earth frame be the same as the camera frame at time $t = 0$.

$$\mathrm{T}_k = \delta\mathrm{T} * \mathrm{T}_{k-1} \qquad (8)$$

where T_k, T_{k-1} and $\delta\mathrm{T}$ are homogeneous transformation matrices at time $t = k$, $t = k-1$ with respect to the earth frame and time $t = k$ with respect to $t = k-1$ respectively.

IV. AGOF/Linear Kalman Filter for IMU-Vision Fusion

In this work, a linear estimator, called "AGOF/Linera Kalman Filter", is designed to integrate visual and inertial measurements together for their complementary properties. A multi-rate fusion method proposed in [19] is chosen to fuse measurements with two kinds of sampling rates. Moreover, the absolute scale factor can be inferred by fusing them together.

A. State Vector Definition

The state vector at time $t = k$ consists of camera position ${}^c_e p_k$ without scale, camera velocity ${}^c_e v_k$, camera acceleration ${}^c_e a_k$ expressed in meters, the reciprocal of the absolute scale factor $\lambda_k = 1/s_k$ which leads to low-order polynomials, the accelerometer bias $b_{a,k}$; so the state vector and the system process noise are expressed as follows:

$$\begin{aligned} x_k &= [{}^c_e p_k; {}^c_e v_k; {}^c_e a_k; \lambda_k; b_{a,k}]; \\ n &= [n_a; n_\lambda; n_{b_a}] \end{aligned}$$

B. Dynamic Model

It is assumed that the system has a uniformly accelerated linear translation at time k and the time span between k and $k+1$ is T. The translation of the camera can be modeled by an equation set. A random walk model is used to estimate λ_k and the biases $b_{a,k}$ based on the value and a white noise at time k. Therefore, the dynamic model of the state is defined as:

$$\begin{aligned} {}^c_e p_{k+1} &= {}^c_e p_k + T\lambda_k {}^c_e v_k + \frac{T^2\lambda_k}{2} {}^c_e a_k + \frac{T^3\lambda_k}{6} n_a \\ {}^c_e v_{k+1} &= {}^c_e v_k + T {}^c_e a_k + \frac{T^2}{2} n_a \\ {}^c_e a_{k+1} &= {}^c_e a_k + T n_a \\ \lambda_{k+1} &= \lambda_k + n_\lambda \\ b_{a,k+1} &= b_{a,k} + T n_{b_a} \end{aligned}$$

C. Measurement Model

Since the two sensors have different sampling rates, two kinds of measurements are considered: ${}^s y^m_k = \left[{}^c_e a^m_k\right]$ for available inertial measurements and ${}^c y^m_k = \left[{}^c_e p^m_k\right]$ for available vision measurements. The measurement update equation for output states is:

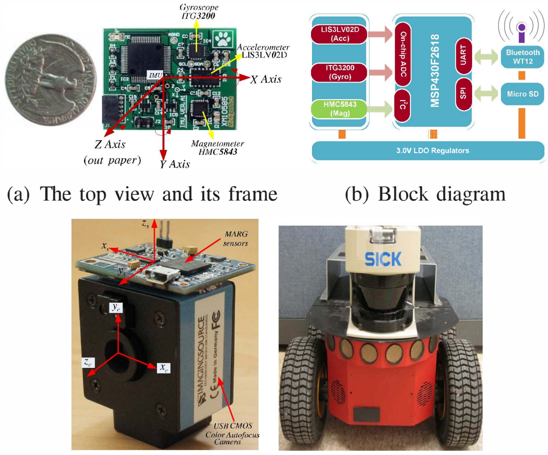
$$y_k \quad = \quad H x_k + e_k \qquad (9)$$

where $H_{s,k} = \begin{pmatrix} 0_{3\times3} & 0_{3\times3} & I_{3\times3} & 0_{3\times4} \end{pmatrix}$ for available inertial measurements or $H_{c,k} = \begin{pmatrix} I_{3\times3} & 0_{3\times3} & 0_{3\times4} & 0_{3\times3} \end{pmatrix}$ for available vision measurements.

The measurement input from the vision part is the camera position without the absolute scale which is calculated by our vision algorithm. Because the acceleration is used in the camera frame, the measurement input models from the inertial sensor shown in (10) is required for converting the raw measurements from the sensor frame into the camera frame and getting rid of the effect of the gravity ${}^e g = [0 \quad 0 \quad -9.8]^T$ expressed in the earth frame.

$$\begin{aligned} {}^c_e a^m_k &= \mathscr{R}({}^c_s q) * (\mathscr{R}({}^s_e \hat{q}_{f,k}) * {}^s a_k - {}^e g) + {}^c_s b \\ {}^s a_k &= {}^s a^m_k - b_{a,k} - e_{a,k} \end{aligned} \qquad (10)$$

where ${}^s a^m_k$ is the raw measurements from the accelerometer in the sensor frame at time k; ${}^c_s q$ and ${}^c_s b$ can be obtained



(a) The top view and its frame (b) Block diagram

(c) vIMU system (d) Pioneer robot

Fig. 5. Prototype and Hardware design for MARG sensor arrays, setup for vIMU system and Pioneer robot platform.

from the hand-eye calibration; the operator $\mathscr{R}$ represents converting orientation from unit quaternion representation to rotation matrix representation.

V. EXPERIMENTAL VALIDATION AND RESULTS

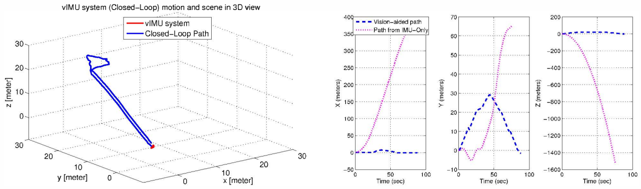
A. Experimental Validation

Figure 5(a) shows the newly developed MARG sensor described in [13]. For the vision input, a USB CMOS color auto-focus camera (DFK 72AUC02-F) from the Imaging Source is used with a resolution of 640×480 and a frame-rate up to $52fps$. In this paper, a low frame-rate $3fps$ is used as the main rate to satisfy epipolar geometry constraints for an accurate baseline and the frame-rate is adaptively adjusted according to our proposed AGOF-aided camera frame-rate adjustment algorithm. The integrated inertial-vision system with the corresponding coordinate frames is depicted in Figure 5(c). The threshold values used in this paper were set as: $\varepsilon = 0.001$, $\varepsilon_\omega = 18°/s$, $\varepsilon_a = 1.5m/s^2$, $\varepsilon_c = 5pixels$.

B. Experimental Results

1) Test I: Straight line movement in X-Y Plane: In this test scenario, a straight line movement was performed in a hallway environment. Figure 6(a) shows the comparative results. It is clearly seen that the result from our proposed method is better than EKF. There is also a magnified area in final location shown in Figure 6(a). The compensated inertial measurements for the adaptive frame-rate adjustment are shown in Figure 6(c), and it is obvious that the robot moves so slowly and smoothly that the frame-rate keeps 3 frame/s in this test.

2) Test II: Curve movement in X-Y Plane: During this test, the Pioneer robot with the sensor rig moved in our office laboratory to demonstrate the accuracy of a curve movement. As shown in Figure 6(d), this test has two severe rotational movements and one translational movement, so the frame-rate of the camera is adaptively adjusted based on the compensated inertial measurements. Figure 6(b) shows the comparative results. It is clearly seen that our proposed method with adaptive frame-rate is the best. There is also a magnified area in final location shown in Figure 6(b). Moreover, the results indicate that the adaptive frame-rate adjustment scheme can improve the accuracy of estimation during severe rotational or translation movement, where the frame-rate is increased from 3 frame/s to 6 frame/s in this test.



(a) Estimated path in 3D view (b) Comparative results in 3 axes

Fig. 7. Estimation results for a closed-loop path with a hand-held vIMU system. There is no ground truth used for this path, but the estimated path should return to the initial position.

3) Test IV: Closed-loop path with hand-held IMU-Vision system: During the test IV, the IMU-Vision system moved through a closed-loop path in 3D view shown in Figure 7(a) while being carried by a person. The total length is about 100 meters. During the first few seconds and the last few seconds of the experiment, the system was kept still. As shown in Figure 7(b), the estimated path from the IMU-only drifts excessively, while the vision system can correct this drift and make the estimated path return to the origin at last.

4) Scale Factor Analysis: Figure 8 shows the scale factor estimation for three different translational movements and illustrates that the scale factor s changes with time t. In these three experiments, different scale factors can be obtained for different experiments and the converge time for the scale factor s is about $10s$. Therefore, $10s$ calibration is required for each experiment before we start each experiment.

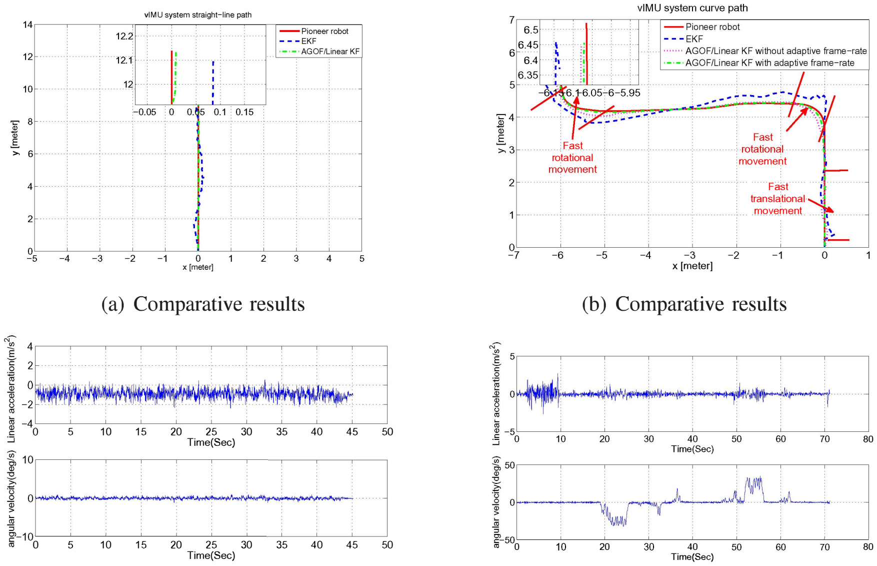
5) Accuracy Analysis: Two different movements, which are straight line, and curve movement, have been used to test the accuracy of our algorithm. The error of each camera position $p_{ec,t}$ in the reconstructed trajectory is calculated as the Euclidean distance between each point of the estimated camera trajectory and the trajectory $p_{robot,t}$ from the Pioneer robot, both in the earth frame. Therefore, this error can be computed as:

$$error_t = \sqrt{(p_{ec,t} - p_{robot,t})^T (p_{ec,t} - p_{robot,t})} \qquad (11)$$

Table I depicts the error accuracy analysis for three experiments. First, the true length of the straight trajectory is 12 meters. Second, the robot moves in the room environment including translation plus orientation where the robot describes a total trajectory of 12.5 meters. Lastly, a semicircle movement is generated by controlling the remote. From Table.I, our proposed method is better than EKF.

VI. CONCLUSIONS

The integration of wearable inertial and visual sensors with an AGOF/Linear Kalman Filter has been proposed for robust human navigation. The use of pre-estimated orientation from MARG sensors can help to eliminate mismatched points based on the properties of the pure camera translation and thus the computation time can be significantly reduced compared to the RANSAC algorithm. In addition, an adaptive-frame-rate single camera is selected to not only avoid motion



(a) Comparative results

(b) Comparative results

(c) Inertial data

(d) Inertial data

Fig. 6. Estimation results for straight line and curve movement in X-Y plane with magnified area in final location.

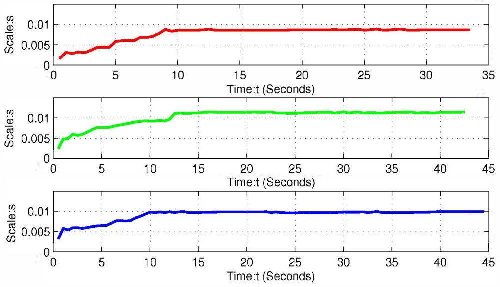


Fig. 8. Scale factor for three different translational movements.

blur based on compensated angular velocity and acceleration but also make an effect called visual zero-velocity update for static motion. Since the accelerometer readings are obtained in absolute metric units, so the metric scale of the monocular camera can be obtained with the help of the inertial sensors, neither using stereo camera or any other calibrated objects.

TABLE I

ERROR ACCURACY ANALYSIS IN TWO EXPERIMENTS

	Trajectory length [m]	Mean error [m]	Maximum error [m]	Mean error over the trajectory [%]
EKF	Test I: 12	0.14	0.275	1.17%
	Test II: 12.5	0.23	0.65	1.84%
KF	Test I: 12	0.064	0.175	0.53%
	Test II: 12.5	0.13	0.35	1.04%

REFERENCES

[1] D. Titterton and J. Weston, *Strapdown Inertial Navigation Technology, Second Edition*. AIAA, 2005.

[2] C. Jekeli, *Inertial Navigation Systems with Geodetic Applications*. De Gruyter, Walter, Inc., Berlin, Germany, 2000.

[3] E. Nebot, S. Sukkarieh, and H. Durrant-Whyte, "Inertial navigation aided with gps information," *Proceedings Fourth Annual Conference on Mechatronics and Machine Vision in Practice*, pp. 169–174, 2006.

[4] A. Schumacher, "Integration of a gps aided strapdown inertial navigation system for land vehicles," *Science*, p. 67, 2006.

[5] I. Skog, *A low-cost GPS aided inertial navigation system for vehicle applications*. PhD thesis, Citeseer, 2005.

[6] P. Corke, J. Lobo, and J. Dias, "An introduction to inertial and visual sensing," *The International Journal of Robotics Research*, vol. 26, no. 6, pp. 519–535, 2007.

[7] P. Gemeiner, P. Einramhof, and M. Vincze, "Simultaneous motion and structure estimation by fusion of inertial and vision data," *The International Journal of Robotics Research*, vol. 26, pp. 591–605, Jan 2007.

[8] D. Randeniya, S. Sarkar, and M. Gunaratne, "Vision-imu integration using a slow-frame-rate monocular vision system in an actual roadway setting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, pp. 256–266, 2010.

[9] L. Armesto, J. Tornero, and M. Vincze, "Fast ego-motion estimation with multi-rate fusion of inertial and vision," *The International Journal of Robotics Research*, vol. 26, no. 6, pp. 577–589, 2007.

[10] Y. Tian, H. Wei, and J. Tan, "An adaptive-gain complementary filter for real-time human motion tracking with marg sensors in free-living environments," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. PP, no. 99, p. 1, 2012.

[11] J. C. K. Chou, "Quaternion kinematic and dynamic differential equations," *IEEE Transactions on Robotics and Automation*, vol. 8, no. 1, pp. 53–64, 1992.

[12] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America A*, vol. 4, no. 4, p. 629, 1987.

[13] S. Hu, X. Chen, and J. Tan, "Pams: a wearable physical activity monitoring system for continuous motion capture in free-living environments," in *Proceedings of the Fifth International Conference on Body Area Networks*, pp. 233–239, 2010.

[14] E. R. Bachmann, *Inertial and magnetic tracking of limb segment orientation for inserting humans into synthetic environments*. PhD thesis, Naval Postgraduate School, 2000.

[15] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. the Press Syndicate of the University of Cambridge, 2003.

[16] J. Lobo and J. Dias, "Vision and inertial sensor cooperation using gravity as a vertical reference," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1597–1608, 2003.

[17] X. Yun, J. Calusdian, E. Bachmann, and R. McGhee, "Estimation of human foot motion during normal walking using inertial and magnetic sensor measurements," *Instrumentation and Measurement, IEEE Transactions on*, vol. 61, no. 7, pp. 2059 –2072, 2012.

[18] B. K. Horn, "Recovering baseline and orientaiton from essential matrix," *Journal of the Optical Society of America*, vol. M, pp. 1–10, 1990.

[19] L. Armesto, J. Tornero, and M. Vincze, "Multi-rate fusion with vision and inertial sensors," 2004.