Monocular Camera and IMU Integration for Indoor Position Estimation

Yinlong Zhang [1] Jindong Tan [2] Ziming Zeng [3] Wei Liang [4,*] Ye Xia [5]

Abstract— **This paper presents a monocular camera (MC) and inertial measurement unit (IMU) integrated approach for indoor position estimation. Unlike the traditional estimation methods, we fix the monocular camera downward to the floor and collect successive frames where textures are orderly distributed and feature points robustly detected, rather than using forward oriented camera in sampling unknown and disordered scenes with pre-determined frame rate and auto-focus metric scale. Meanwhile, camera adopts the constant metric scale and adaptive frame rate determined by IMU data. Furthermore, the corresponding distinctive image feature point matching approaches are employed for visual localizing, i.e., optical flow for fast motion mode; Canny Edge Detector & Harris Feature Point Detector & Sift Descriptor for slow motion mode. For superfast motion and abrupt rotation where images from camera are blurred and unusable, the Extended Kalman Filter is exploited to estimate IMU outputs and to derive the corresponding trajectory. Experimental results validate that our proposed method is effective and accurate in indoor positioning. Since our system is computationally efficient and in compact size, it's well suited for visually impaired people indoor navigation and wheelchaired people indoor localization.**

I. INTRODUCTION

Recently, indoor position estimation is increasingly showing itself in a variety of areas such as the visually impaired navigation assistance, wheelchaired people indoor localizing, pedestrian shopping guidance in department stores, traditional robot autonomous navigation, simultaneous localization and mapping (SLAM) [1,2]. The current available indoor positioning methods include inertial positioning, mobile camera positioning, cellphone network positioning, light sensor positioning and Wi-Fi positioning [3]. Among these methods, inertial sensors, integrated with 3-axis accelerometer, 3-axis gyroscope and 3-axis magnetometer, are characteristic of compact size, highly dynamic response, low cost and portability, which displays the potentials and competitiveness for indoor positioning. However, this type of sensor merely works over short periods of time due in large part to its inconstant errors (e.g., biases error, scale factor error), needlessly to say the acceleration double integrations [4]. Yet, it's remarkable that Foxlin et al. proposed to mount an IMU rigidly on foot and adopted Zero-Velocity-Update (ZUPT) [5] for indoor positioning which proves that wearable sensors are capable of obtaining the potentially accurate position and in turn dramatically decrease acceleration drifts. But orientation estimate (yaw angle) from gyroscope is still not accurate enough in the long term. By comparison, monocular camera features in accuracy and effectiveness for indoor position estimation, especially in the areas of Kinect-SLAM for unstructured environment [6], Robot navigation [7] for it's capable of providing crucial measurements such as translation and rotation by tracking features between successive frames. However, the camera predefined frame rate fails to satisfy dynamic requirements and thus poses challenges to real-time positioning. The recently proposed integration of camera and IMU (V-IMU) is able to improve the overall positioning performance in the sense that they are complementary in accuracy and frequency response [8]. Yet, the previous literatures on V-IMU system inherently suffers from several drawbacks [4,9,10]: the camera being fixed forward to collect unknown and unstructured scenes part of which the objects captured in images are also moving; to deduce the inaccurate metric scale in each frame which causes miscalculations in position estimation; V-IMU system sets the constant frame rate for camera which results in image blurring with low sampling rate for fast motion and camera oversampling for slow motion. With respect to these inertial and visual problems, we propose a novel Monocular Camera and IMU (MC-IMU) integration system for indoor positioning. Our system is capable of recognizing the moving patterns and altering the camera frame rate accordingly. The corresponding feature point matching algorithms is also determined via IMU outputs. In addition, the system is able to send the static-state feedback detected by camera to IMU and conduct the IMU drift correction. The proposed strategy effectively improves the image processing computational efficiency (avoids fast-motion image blurring, slow-motion camera oversampling) and reduces IMU long-term accumulated drift. It's remarkable that our camera is being fixed downward collecting the floor scenes of which textures are orderly distributed and feature points may well be robustly detected. This indoor positioning system is literally low cost and energy efficient which can be applied to scenarios like assisting visually impaired people indoor navigation, wheelchaired people indoor positioning and pedestrian shopping in department store.

*This work was supported by the National Science Foundation of China under contact (61172145, 61333019)

[1] Yinlong Zhang is with the Key Laboratory of Networked Control Systems, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, 110016, China and with University of Chinese Academy of Sciences, Beijing, 100049, China. zhangyinlong@sia.cn

[2] Jindong Tan is with the Department of Mechanical, Aerospace and Biomedical Engineering, University of Tennessee, Konxville, TN 37996,USA. tan@utk.edu

[3] Ziming Zeng is with the Information and Control Engineering Faculty, Shenyang Jianzhu University, Shenyang, 110168, China zengziming1983@gmail.com

[4,*] Wei Liang (Corresponding author) is with the Key Laboratory of Networked Control Systems, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, 110016, China. weiliang@sia.cn

[5] Ye Xia is with the Key Laboratory of Networked Control Systems, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, 110016, China. xiaye@sia.cn

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

II. METHODOLOGY

In this paper, the general idea is to achieve the monocular camera and IMU integration for indoor positioning that is suitable for discrepant moving patterns: static motion, slow motion, fast motion and superfast motion among which visual positioning is operated upon slow motion and fast motion; IMU positioning is conducted upon superfast motion and IMU drift correction is performed upon static motion. The overall architecture of our method is shown in (Fig. 1). To achieve this goal, the camera sampling frame rate and the corresponding image feature matching algorithms are adaptively determined by IMU sensor data. Meanwhile, the camera position estimation is, in turn, able to feedback the static state to IMU for drift correction under static motion conditions. Finally, the IMU positioning is applied for superfast motion.

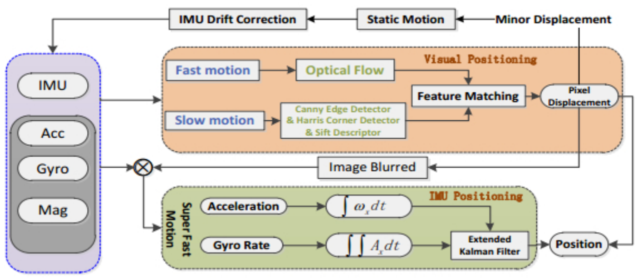


Fig. 1. The flowchart of the proposed method for indoor positioning

A. Visual Positioning

Due to the fact that camera sampling rate stays constant, it fails to suffice the moving object dynamic needs. Thus, we set two sampling frame rates for camera and adopt distinctive feature extracting and matching algorithms respectively.

1) Fast Motion Positioning: For estimating the mobile camera fast motion, we have set the higher camera frame rate mode. Higher frame rate is selected if outputs from IMU (three axis accelerations: $\overrightarrow{\mathbf{A}}$, three-axis gyro rates: $\overrightarrow{\mathbf{W}}$ are larger than the pre-defined values $\overrightarrow{\delta_a}$, $\overrightarrow{\delta_\omega}$ where $\overrightarrow{\delta_a}$ denote the thresholds for 3-axis accelerations and $\overrightarrow{\delta_\omega}$ denote the thresholds for 3-axis turn rates. Afterwards, key points in images are firstly detected using Harris detector. Then we apply computationally efficient tracking algorithm: sparse optical flow via a series of frames.

Optical flow is based upon two assumptions that corresponding pixel values between successive images have constant brightness over time and that nearby points in the image move in a similar manner.

- Pixel value constancy assumption

$$I(x,y,t) = I(x+\Delta x, y+\Delta y, t+\Delta t) \tag{1}$$

where (x, y) denotes the key point pixel position and I denotes the pixel grey value at time instant t. After one-step Taylor expansion over t, the optical flow constraints are yielded as below:

$$I_x u_x + I_y u_y + I_t = 0 \tag{2}$$

In the above formula, Ix and Iy mean the gradients in X and Y directions; It represents the pixel change over time.

- Nearby points similar motion assumption

By means of employing the adjacent pixels around the feature point, we could obtain a set of equations shown in equation (3). Ultimately, the optical velocity $[u_x, u_y]^T$ will be attained via least-square solutions.

$$\begin{bmatrix} I_{x1} & I_{y1} \\ I_{x2} & I_{y2} \\ \dots & \dots \\ I_{xn} & I_{yn} \end{bmatrix} \begin{bmatrix} u_x \\ u_y \end{bmatrix} = \begin{bmatrix} I_{t1} \\ I_{t2} \\ \dots \\ I_{tn} \end{bmatrix} \tag{3}$$

2) Slow Motion Positioning: For slow motion indoor positioning, we have adopted the lower sampling frame rate and an accurate algorithm for key point matching. Unlike the fast motion mode that requires computational efficiency and dynamic response, the accuracy and robustness for key point matching takes priority over computational load since longer image processing time is permitted in this motion mode. As for our intentionally designed downward fixed mobile camera, the floors are characteristic of orderly distributed perpendicular lines which offer more manifest key point (corners) to track. The slow motion positioning flow chart is shown in Fig. 2.

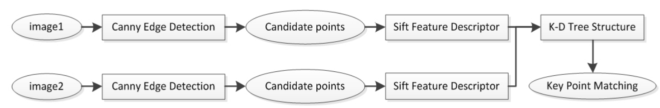


Fig. 2. Flow chart for feature point matching in slow motion mode

Thus, we directly harness canny edge detector to select the candidate points among which Harris detector (that takes uniquely larger Hessian-matrix eigenvalues into account) is utilized in determining the feature points. Afterwards, sift descriptor, which employs a normalized vector consisting of 128 elements, is applied to represent the key point. Finally, we use K-D tree (an optimized data structure for nearest neighbor searching) for key point matching between adjacent images.

B. IMU Positioning and Drift Correction

1) IMU Positioning: During indoor position estimation, what invariably happens is that the images from mobile camera will be blurred or corrupted, due to body frame fast rotation or high velocity. In this case, we have to resort to make use of IMU for instantaneous estimation for it is relatively quick to dynamic responses. As indoor environment occasionally suffers from magnetic disturbances, we merely use 3-axis accelerations and 3-axis gyro rates. After gravitational acceleration compensation, double integration of three axis accelerations: $\overrightarrow{A}$ multiplied by integration of three-axis gyro rates: $\overrightarrow{W}$ will yield instantaneous displacement changes in X and Y directions.

The Extended Kalman Filter is used to estimate the instantaneous body frame positions since the estimated states are expressed in nonlinear formulations. The state vector here is defined as

$$X = [{}^sP_t^T, {}^s\upsilon_t^T, {}^s\alpha_t^T, {}^s\theta_t^T, {}^s\omega_t^T, {}^sq_t^T, {}^s_b\alpha_t^T, {}^s_b\omega_t^T]^T \tag{4}$$

${}^sP_t = [{}^sP_{x,t}, {}^sP_{y,t}, {}^sP_{z,t}]^T$: vector of IMU 3-axis positions;
${}^s\upsilon_t^T = [{}^s\upsilon_{x,t}, {}^s\upsilon_{y,t}, {}^s\upsilon_{z,t}]^T$: vector of IMU 3-axis velocities;
${}^s\alpha_t = [{}^s\alpha_{x,t}, {}^s\alpha_{y,t}, {}^s\alpha_{z,t}]^T$: vector of IMU 3-axis accelerations;
${}^s\theta_t = [{}^s\theta_{x,t}, {}^s\theta_{y,t}, {}^s\theta_{z,t}]^T$: vector of IMU 3-axis euler angles;
${}^s\omega_t = [{}^s\omega_{x,t}, {}^s\omega_{y,t}, {}^s\omega_{z,t}]^T$: vector of IMU angular rates;
${}^sq_t = [{}^sq_{w,t}, {}^sq_{x,t}, {}^sq_{y,t}, {}^sq_{z,t}]^T$: vector of IMU quaternions;
${}^s_b\alpha_t = [{}^s_b\alpha_{x,t}, {}^s_b\alpha_{y,t}, {}^s_b\alpha_{z,t}]^T$: bias errors for accelerations;
${}^s_b\omega_t = [{}^s_b\omega_{x,t}, {}^s_b\omega_{y,t}, {}^s_b\omega_{z,t}]^T$: bias errors for angular rates;

All the vectors are defined in the sensor frame at time instant t. Given the above state vector, the dynamic model is defined as follows:

$${}^sP_t = {}^sP_{t-1} + \Delta t \cdot v_{t-1} + \frac{\Delta t^2}{2} \cdot {}^s a_t + \frac{\Delta t^3}{6} \cdot {}^s_b a_t \tag{5}$$

$${}^sv_t = {}^sv_{t-1} + \Delta t \cdot {}^s\alpha_t + \frac{\Delta t^2}{2} \cdot {}^s_b\alpha^t \tag{6}$$

$${}^s\alpha_t = {}^s\alpha_{t-1} + \Delta t \cdot {}^s_b\alpha^t \tag{7}$$

$${}^s_b\alpha_t = {}^s_b\alpha_{t-1} + \Delta t \cdot b_a \tag{8}$$

$${}^s\theta_t = \begin{bmatrix} arctan(2({}^sq_{w,t} \cdot {}^sq_{x,t} + {}^sq_{y,t} \cdot {}^sq_{z,t}), 1 - 2({}^sq_{y,t}^2 + {}^sq_{x,t}^2)) \\ arcsin(2({}^sq_{w,t} \cdot {}^sq_{y,t} - {}^sq_{x,t} \cdot {}^sq_{z,t})) \\ arctan(2({}^sq_{w,t} \cdot {}^sq_{z,t} + {}^sq_{x,t} \cdot {}^sq_{y,t}), 1 - 2({}^sq_{y,t}^2 + {}^sq_{x,t}^2)) \end{bmatrix} \tag{9}$$

$${}^sq_t = {}^sq_{t-1} + \frac{1}{2} \cdot ({}^sq_{t-1} \bigotimes \omega_t^T) \tag{10}$$

$${}^s_b\omega_t = {}^s_b\omega_{t-1} + \Delta t \cdot \omega_a \tag{11}$$

Function (5)∼(8) are related to body frame positions; Function (9)∼(11) correspond to body frame orientations. In (8) and (11), b_a and ω_a are white noises that satisfy zero-mean Gaussian distribution.

The measurement model is $y_t = {}^sHx_t + {}^se_t$. Since the observed values are merely accelerations, gyro rates and quaternions, the observation matrix is:

$${}^sH = [O_{3\times3}, O_{3\times3}, I_{3\times3}, O_{3\times3}, I_{3\times3}, I_{4\times4}, O_{3\times3}, O_{3\times3}] \tag{12}$$

where $0_{m\times m}$ denotes $m \times m$ square matrix with all elements zero values; $I_{n\times n}$ indicates $n \times n$ identity matrix.

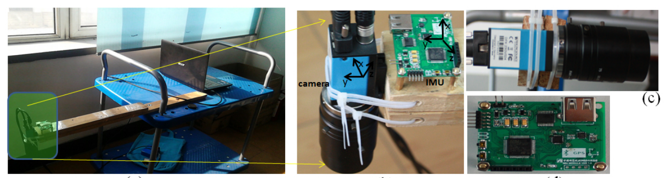
2) IMU Drift Correction: During monocular camera tracking, if the feature point pixel changes subtly between consecutive frames for a period of time, we have every reason to believe that the object is in the stage of static motion. However, the inertial readings from IMU, for most of the cases, turn out to be nonzero values. Then we propose to enable the vision system to feedback this static state to IMU for conducting drift correction. To the best of our knowledge, this is a very effective solution for IMU to eliminate the accumulated errors from historical inertial data.

III. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we introduce the MC-IMU system hardware design, error analysis and the positioning experiments that include short-term positioning and long-term positioning.

A. MC-IMU Hardware Design

In our indoor positioning system, the monocular camera and IMU board are fixed in alignment onto the cart, see Fig. 3. The camera (Imaging Source product: DFK23GV024), samples the data at 15fps for slow motion mode and 60fps for fast motion mode, both with a resolution of 752×480 pixels. The constant focal length is set as 7.5mm. IMU board is designed with an inertial module-MPU6500 integrated with a MEMS three-axis accelerometer (full scale range16g), three-axis gyroscope (full scale range ±2000 degree/sec) and three-axis magnetometer. The IMU board size is 52mm40mm14mm. The inertial signals are sampled at 100Hz and interfaced to computer by Bluetooth.



(a) (b) (c) (d)

Fig. 3. Monocular camera and IMU for Indoor Positioning System.(a) Monocular camera and IMU fixed onto the cart; (b) Monocular camera and IMU; (c) Monocular camera and lens; (d) Our Designed IMU Board

B. Error Analysis

The absolute error e_a and the average error E_a are introduced to measure the system performance as below:

$$e_a = \sqrt{x_e^2 + y_e^2} \qquad E_a = \frac{1}{n}\sum_{i=1}^{n} e_{a,i} \tag{13}$$

where x_e and y_e are position errors in X and Y directions, n is the number of experiments within the same group.

In addition, we have developed a new error analysis parameter: the time weighted average error E_{wa}, which is shown as:

$$E_{wa} = \sum_{i=1}^{n} E_{a,i} \frac{T_i}{T_{all}} \tag{14}$$

where T_i denotes the periods of time for each test which is added together as T_{all}, i.e., $T_{all} = \sum_{i=1}^{m} T_i$. This novelly introduced parameter E_{wa} is capable of reflecting the distinctive monocular camera adaptive frame rates and the corresponding feature points matching algorithms on the positioning error analysis. In determining the body frame motion modes, i.e., slow motion mode and fast motion mode, we set the IMU accelerations thresholds $\overrightarrow{\delta}_a = [\delta_{a,x}, \delta_{a,y}, \delta_{a,z}]$ and gyro rates thresholds $\overrightarrow{\delta}_\omega = [\delta_{\omega,x}, \delta_{\omega,y}, \delta_{\omega,z}]$. In the experiments, we set $\|\overrightarrow{\delta}_a\| = 1.8g, \|\overrightarrow{\delta}_\omega\| = 1.6rad/s$

C. Positioning Results and Analysis

1) Short-term Positioning: For the short-term positioning, we develop two experiments: straight-line positioning and staircase positioning.

- Straight Line Positioning

A specific distance of 35m is measured. The starting point and ending point are marked with black and green dot precisely. The blue dashed line is the pre-determined path and the red one is the reconstruction trajectory. The experiments are performed 5 times and the average error is 0.42

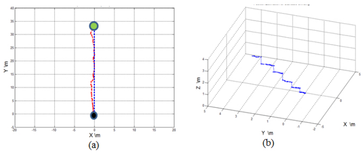


Fig. 4. Short-term positioning trajectory.(a) Line Trajectory; (b) Staircase trajectory

- Staircase Positioning

The subject holds the monocular camera and IMU which are aligned and wired together. The monocular camera measures the subject displacements in X-Y plane. In this test, IMU not only determines the subject motion mode for camera frame-rate determination but also measures the position changes in Z direction. The total staircase is 2 meters in height and 3.5 meters in length. Two sets of experiments are conducted. Each experiment in the first group is completed within 10 seconds. The other experiment in the second group is finished within 30 seconds. As is shown in TABLE 1, the time weighted error is 0.107m and the relative error is 0.658%, both of which are in high accuracy.

TABLE I

ERROR ANALYSIS OVER STAIRCASE TRAJECTORY

Group	Time(s)	$e_a(m)$	$E_a(m)$	$E_{wa}(m)$	$E_r(\%)$
1	9.5	0.13	0.19	0.107	0.658
	8.9	0.21			
	10.4	0.23			
2	28.1	0.106	0.0787		
	29.5	0.055			
	27.6	0.075			

2) Long-term Positioning: Rectangle-shaped path is designed with the total length approximated 210m. The reconstruction result is shown in Fig. 5.

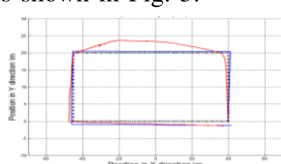


Fig. 5. Long-term Positioning Trajectory

The black dashed line is the predefined rectangle path; the blue one is the reconstruction trajectory via our proposed method; the red one is the outcome of using merely the monocular camera with the constant low frame rate and the constant feature matching algorithm.

It is apparent that the trajectory results from our proposed method, as shown in TABLE 2 group 1, greatly improves the overall positioning performance, which compares favorably to the constant image matching algorithm in TABLE 2-group 2-Test 2 [10] that is not accurate due to its constant sampling rate to dynamic responses, which is not appropriate for dynamic moving patterns. In addition, our trajectory reconstruction computational time is approximately 580 seconds which are quite less than the method group 2-Test 1 [5] that takes 687 seconds in video sequence feature point matching, which is relatively operational complicated.

TABLE II

ERROR ANALYSIS OVER RECTANGLE TRAJECTORY

Group	Trajectory(m)	$e_a(m)$	Time(s)	$E_r(\%)$
1	Test1: 210	3.01	573	1.42
	Test2: 212	2.61	584	1.23
2	Test1: 215 [4]	5.27	687	2.45
	Test2: 208 [9]	5.81	632	2.79

IV. CONCLUSION AND FUTURE WORK

In this paper, we present an implementation of visual and inertial integration approach for indoor position estimation. There are three advantages of our MC-IMU system which are listed as follows: Firstly, the camera frame rate and corresponding feature point matching algorithms are adaptively determined: higher sampling rate and the computationally efficient optical flow algorithm are adopted for fast motion that are reflected by higher IMU outputs; lower sampling rate and the accurate image feature point matching algorithm-Canny Edge detector & Harris Feature point detector & Sift descriptor are employed for slow motion that are reflected by lower IMU readings. Secondly, we resort to IMU positioning based upon extended Kalman filter for the short-term superfast and abrupt motion where the camera fails to capture the clear picture. Ultimately, the camera position estimation is in turn, able to feedback the static state to IMU for IMU drift correction under static motion conditions. In the future, in-depth multi-rate fusion between IMU and camera will be investigated for the better and precise indoor positioning.

REFERENCES

[1] L.A. Guerrero, F. Vasquez and S.F. Ochoa, ”An indoor navigation system for the visually impaired,” *Sensors*, vol.37, no.3, pp. 1687-1499, Jan. 2013.

[2] J.A.B. Link, F. Gerdsmeier, P. Smith and k. Wehrle, ”Indoor navigation on wheels (and on foot) using smartphones,” in *Proc. 4th Annu. Indoor Positioning and Indoor Navigation,* Sydney, 2012, pp.1-10.

[3] R. Mautz, ”Indoor Positioning Technologies,” Ph.D.dissertation, Dept. Civil, Environmental and Geomatic. Eng,. ETH Zurich., Switzerland, 2012.

[4] Y. Tian, J. Zhang and J. Tan, ”Adaptive-frame-rate monocular vision and IMU fusion for robust indoor positioning,” in *Porc. IEEE International Conf. Robotics and Automation (ICRA)*, Karlsruhe, Germany, 2013, pp. 2257-2262.

[5] X. Chen, S. Hu, Z. Shao and J. Tan, ”Pedestrian positioning with physical activity classification for indoors,” in *Proc. International Conf. Robotics and Automations (ICRA)*, Shanghai, 2011, pp. 1311-1316.

[6] J. Hartmann, D. Forouher, M. Litza, J.H.Klussendorff and E. Maehle, ”Real-time visual SLAM using FastSLAM and the Microsoft kinect camera,” in *Proc. 7th Annu. ROBOTIK*, Munich, 2012, pp. 458-463.

[7] J. Biswas and M. Veloso, ”WiFi Localization and Navigation for Autonomous Indoor Mobile Robots,” in *Proc. International Conf. Robotics and Automations(ICRA),* Anchorage, 2010, pp. 458-463.

[8] F.M.Mirzaei and S.I. Roumeliotis, ”A Kalman Filter-Based Algorithm for IMU-Camera Calibration: Observability Analysis and Performance Evaluation,” *IEEE Trans. Robotics.*, vol.24, pp.1143-1155, Oct. 2008.

[9] G. Panahandeh and M. Jansson, ”Vision-Aided Inertial Navigation Based on Ground Plane Feature Detection,” *IEEE/ASME Trans. Mechatronics.*, to be published.

[10] D.I.B. Randeniya, S. Sarkar and M. Gunaratne, ”Vision-IMU Integration Using a Slow-Frame-Rate Monocular Vision System in an Actual Roadway Setting.,” *IEEE Trans. Intelligent Transportation Systems.*, vol.11, pp.256-266, June, 2010.