COMBINING MEMS-BASED IMU DATA AND VISION-BASED TRAJECTORY ESTIMATION

F. Tsai[a*], H. Chang[b] and A. Y. S. Su[c]

[a]Center for Space and Remote Sensing Research
[b]Department of Civil Engineering
[c]Research Center for Advanced Science and Technology
National Central University, Zhong-li, Taoyuan 320 Taiwan
ftsai@csrsr.ncu.edu.tw; 1984chang@gmail.com; ncuaddison@gmail.com

KEY WORDS: Positioning and Tracking, Vision-based trajectory, Motion estimation, IMU, Data fusion

ABSTRACT:

This paper presents an efficient location tracking algorithm that integrates vision-based motion estimation and IMU data. Orientation and translation parameters of the mobile device are estimated from video frames or highly overlapped image sequences acquired with built-in cameras of mobile devices. IMU data are used to maintain continuity of the orientation estimation between sampling of the image homography calculation. The developed algorithm consists of six primary steps: (1) pre-processing; (2) feature points detection and matching; (3) homography calculation; (4) control points detection and registration; (5) motion estimation and filtering; (6) IMU data integration. The pre-processing of the input video frames or images is to control the sampling rate and image resolution in order to increase the computing efficiency. The overlap rate between selected frames is designed to remain above 60% for matching. After pre-processing, feature points will be extracted and matched between adjacent frames as the conjugate points. A perspective homography is constructed and used to map one image to another if the co-planar feature points between subsequent images are fully matched. The homography matrix can provide the camera orientation and translation parameters according to the conjugate pairs. An area-based image-matching method is employed to recognize landmarks as reference nodes (RNs). In addition, a filtering mechanism is proposed to ensure the rotation angle was correctly recorded and to increase the tracking accuracy. Comparisons of the trajectory results with different combinations among vision-based motion estimation, filtering mechanism and IMU data integration are evaluated thoroughly and the accuracy is validated with on-site measurement data. Experimental results indicate that the develop algorithm can effectively estimate the trajectory of moving mobile devices and can be used as a cost-effective alternative for LBS device both in outdoor and indoor environment.

1. INTRODUCTION

Geographically localization and orientation estimation of the mobile devices is an essential component for Location-Based Service (LBS) applications. Nevertheless, the development of effective methods for data integration, interoperability and optimization for LBS is an emerging issue and challenging task as new generations of positioning and tracking devices have become ubiquitous. Due to the unavailability of Global Navigation Satellite Systems (GNSS) signals for indoor positioning, approaches utilizing different sensor types were presented, such as infrared (IR), Radio Frequency IDentification (RFID), Near Field Communication (NFC), ultrasound, Wireless Local Area Network (WLAN), Bluetooth, sensor networks, ultra-wideband, magnetic signals, vision analysis and audible sound (Ching et al., 2013; Gu et al., 2009; Hightower and Borriello, 2001). Each sensor or developed system takes advantage of a particular positioning technology. Combining some of these technologies may improve the effectiveness positioning or tracking, but may also inherit their limitations.

Low-cost Micro Electrical Mechanical System (MEMS) based Inertial Measuring Unit (IMU) and build-in digital cameras are widely implemented in smart phones, Personal Digital Assistant (PDA) and other mobile devices. MEMS-based IMU can provide reliable orientation data in a short period of time. On the other hand, vision-based motion estimation can locate and track the device from the images acquired with the built-in camera. Therefore, combining both should be able to improve the positioning and tracking performance of mobile devices for location-based service applications.

2. RELATED WORKS

Simultaneous Localization And Mapping (SLAM) is one of the applications in Mobile Mapping System (MMS). Monocular SLAM is more challenging but convenient to install and obtain data (Achtelik et al., 2011). This type of device usually has a single camera and can take video or photos with a low cost MEMS sensor providing the equipment's real-time attitudes and acceleration information. They can be used as references for outdoor GNSS or indoor Wi-Fi positioning to improve the accuracy. Time-Of-Flight (TOF) cameras is another technology for ego-motion determination in indoor environments. TOF cameras consist of compact, solid-state sensors that provide depth and reflectance measurements at high frame rates of up to 50 Hz independent from surrounding light, so they can operate very quickly for three-dimensional (3D) imaging (Iddan and Yahav, 2001; Kohoutek et al., 2013).

For mobile devices with built-in (low-cost) digital cameras and MEMS-based IMU, the relative orientation and translation parameters can be estimated from video frames or highly overlapped images acquired using the built-in camera (Chang and Tsai, 2013). On the other hand, MEMS-based IMU can provide reliable absolute orientation data in a short period of time. The positioning system using MEMS-based IMU data is a self-contained navigation technique. Measurements provided by accelerometers and gyroscopes are used to track the position and orientation of an object relative to a known starting point, orientation and velocity (Woodman, 2007).

Multi-sensor based navigation systems can be synchronized for

positioning and tracking. Conventionally, when GNSS signals are not available, trajectory is improperly treated by IMU and Kalman filter (Fakharian et al., 2011). Kelly et al. (2008) proposed an Unmanned Aerial Vehicle (UAV) navigation system which combines stereo visual odometry with IMU. This approach fuses the motion estimates from both sensors in an extend Kalman filter (EKF) to determine vehicle position and attitude. IMU usually has a shifting effect in time. A possible solution is to integrate with vision-assisted methods (Hide et al., 2010). Single camera has a limitation when measuring metric scale, so mapping monocular images must assume the scale is an undetermined factor. Scale factor estimation can be achieved by adding additional information source such as a calibration object (Davison et al., 2007) or by exploiting nonholonomic motion constraints (Scaramuzza et al., 2009). Strasdat et al. (2010) presented an algorithm based on non-linear optimization which can deal with drift effects in large scale scenes. Using a single camera to obtain image features and to trace the feature by EKF can also achieve tracking in unknown scenes to rebuild the path of camera movement (Tsai et al., 2013).

3. MATERIALS AND METHOD

This research focuses on developing a location tracking algorithm that combines vision-based trajectory estimation and low-cost MEMS-based IMU data for hand-held consumer mobile devices. The data used are the images captured by the built-in camera and orientation parameters derived from the MEMS-based IMU. The general procedure of the developed algorithm is illustrated in Figure 1.
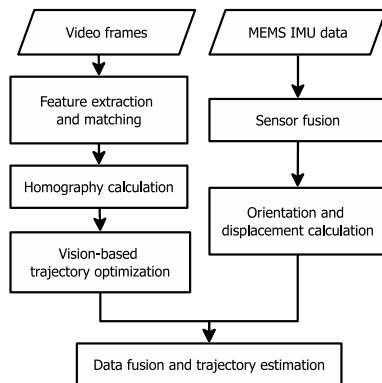

Figure 1: Procedure of the proposed positioning scheme.

3.1 Feature extraction and matching

The first step of vision-based trajectory estimation is to extract and match conjugate points from image frames. Features points are extracted using the Accelerated Segment Test (FAST) corner detector, which uses a circle of surrounding pixels to classify whether a candidate point is a corner pixel or not, based on the intensity differences of contiguous pixels. As described in Eq. (1), if a set of N contiguous pixels in the circle, $I(x)$ are all brighter or darker than the intensity of candidate pixel p (donated by $I(p)$) with a threshold of ε_d, then p is classified as corner.

$$N = \sum_{x \forall (circle(p))} |I(x) - I(p)| < \varepsilon_d \qquad (1)$$

After extracting feature points between subsequent images, Speeded-Up Robust Features (SURF) is then used to find blob features for each point. SURF descriptor uses wavelet responses in horizontal and vertical directions. The advantages of SURF are scale and rotation invariant, resistance of noise in position and intensity, and high match precision. The formula of each sub-region with a four-dimensional descriptor vector, v, is shown in Eq. (2). All 4 by 4 sub-regions can be concatenated as a descriptor vector of length 64.

$$v = \left(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y| \right) \qquad (2)$$

The conjugate points between subsequent images can be matched according to this 64-dimensional descriptor. This study uses Sum of Absolute Differences (SAD) to quickly find correspondences of the feature points. These differences are collected to create a simple metric for similarity check.

3.2 Perspective homography

A perspective homography matrix (H) can be used to map one image to another if the co-planar feature points between subsequent images are fully matched. The projective transformation is shown in Eq. (3). where H can be divided into the metric part (M) and non-metric part (N). M is used for the calibration of orthogonality and parallelism with projective transformation parameters l_1 and l_2. Skewness and aspect ratio are calibrated by affine transformation parameters α and β. The metric part comprises of a 2x2 rotation matrix (R), a 2x1 translation vector (t), and an isotropic scaling (s). Camera motion vector can be derived from t and the orientation is determined from the rotation and scaling factors as described in Eq. (4) and (5). However, outliers such as features extracted from shadows or noises should not be included in the homography calculation.

$$H = MN; M = \begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix}; N = \begin{bmatrix} \frac{1}{\beta} & -\frac{\alpha}{\beta} & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & 1 \end{bmatrix} \qquad (3)$$

$$\theta = \left(\tan^{-1}\left(\frac{R(1,1)}{R(1,2)} \right) + \tan^{-1}\left(\frac{R(2,1)}{R(2,2)} \right) \right)/2 \qquad (4)$$

$$s = \frac{R(1,1) + R(2,2)}{2\cos\theta} \qquad (5)$$

3.3 Trajectory optimization

To recognize the reference nodes (RNs) as control points along the path, this study combines an area-based pattern recognition method with identifications of RNs for calibrating the location estimation. The Normalized Cross-Correlation (NCC) scheme is widely used in image-processing applications, and it is invariant to the brightness difference between the image and template. However, in most of the scenes, there are no RNs available for calibration. Therefore, a self-defined filtering mechanism should be applied for the process. If there is any blunder during the accumulation of the trajectory, the error will propagate. The next is the camera motion drift caused by video vibration or hand-shake effect, where the error is due to the estimation of rotation angle. This error is not significant in the short time period, but will accumulate after error propagation. Therefore, they need to be minimized. Two thresholds are defined in this filtering mechanism. One is checking the continuity of rotating frames so the volatile rotation will be recognized as a noise and removed. The other is the accumulation of the continually rotating angle. This filtering mechanism not only can ensure the rotation angle is correctly recorded, but also increases the tracking accuracy.

3.4 MEMS-based IMU data analysis

Low-cost MEMS-based IMU provides high-frequency and continuous accelerometers, gyroscopes and magnetometers data of the mounted device. The general procedure of the proposed positioning scheme according to MEMS-based IMU data is a sequential process as outlined in Algorithm-1. MEMS-based IMU is unstable and easily interfered by surrounding magnetic field, which makes the filtering mechanism a necessity after data collection. Collected IMU data are then transferred from traditional mechanical gyroscopic values to the attitude and heading reference system (AHRS) for orientation estimation. After that, the translational accelerations can be calculated by rotating the body frame to earth frame according to the estimated orientation. Finally, velocity and the position of the mobile device can be estimate by integrating time and accelerations.

**Algorithm 1** Proposed MEMS-based IMU positioning scheme

---

1. MEMS IMU data input: accelerometer, gyroscope, and magnetometer

2. Data filtering (low-pass and high-pass filters)

3. MARG (Magnetic, Angular Rate, and Gravity) to AHRS (Altitude and Heading Reference System) conversion

4. Rotate body frame to Earth frame (for translational acceleration calculation)

5. Linear acceleration and time integration (for velocity and position estimation)

Although the filtering mechanism is helpful for correcting the errors during the data collection, the drift error can still significantly decrease the accuracy of the IMU-only positioning scheme and it may become completely unusable after a short period of time. In the proposed IMU/camera integrated system, the low quality of vision-based trajectory will be replaced by IMU data, but control points identified in the vision-based trajectory can also be used to correct the IMU errors.
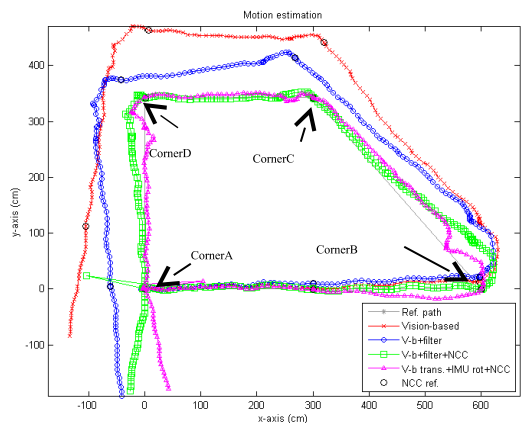
4. EXPERIMENTAL RESULTS

Two data sets were collected and tested in this study, which have 4 and 5 RNs (star-shape marks in Figure 2), respectively. The data were acquired using a consumer smart phone with built-in digital camera and MEMS-based IMU and hand-held by a person while walking along the planed routes. All feature points were also measured by ground-based measurement to be used as ground-truth data for validation. The complete path is a closure loop starting from corner A and walking straight to B, C and D counter-clockwise and finally returning and passing over corner A (grey line in Figure 2). The loop passes through 4 and 5 RNs in a distance of 12.8 and 20.42 meters in the two cases respectively. The video frame resolution is 720x1280 pixels with fixed focal length. Video frames are downsampled to 5 frames per second and at least 60% of overlaping is maintained between adjacent image frames. Collected MEMS-based IMU data and their calculated values of the acceleration, magnetic field, angular velocity and the attitude of the mobile device are illustrated in Figure 3 and 4, respectively.

The red cross markers in Figure 2 demonstrate the estimated trajectory using vision-based methods and fusion of images and IMU data. Filtering mechanism has improved the accuracy by
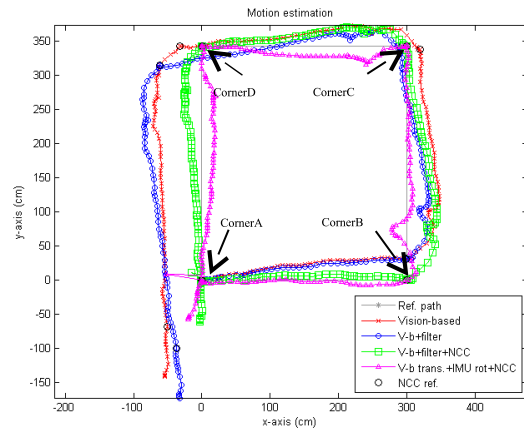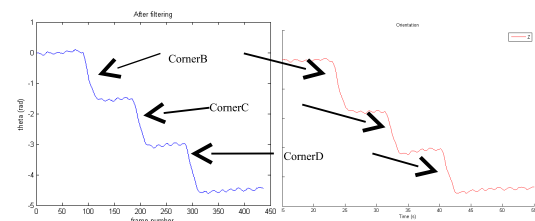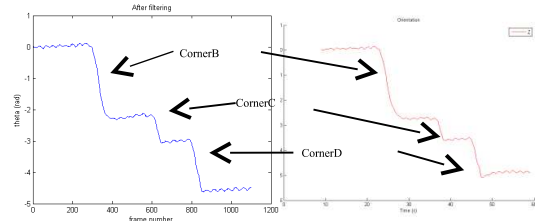

(a) Case 1

(b) Case 2

Figure 2: Experimental results of trajectory estimation. Gray: RNs distribution and walking path; Red: vision-based trajectory estimation; Blue: with filtering mechanism; Green: referenced by NCC matched features; Magenta: estimation from fused image and IMU data.
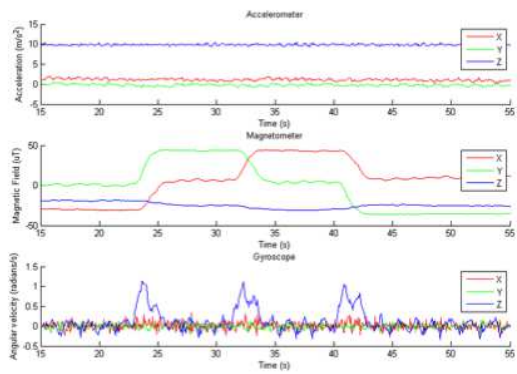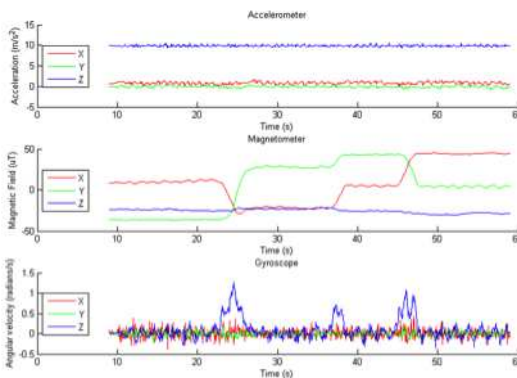

(a) Case 1

(b) Case 1

Figure 3: Rotation angles derived from the accumulation between adjacent sample frames using vision-based method (left) and MEMS-based IMU data (right). The step-shaped features identify the detection of corner frames.

(a) Case 1

(b) Case 2

Figure 4: Collected MEMS-based IMU data; from top to bottom: accelerometer, magnetometer, and gyroscope.

decreasing the errors of rotation estimation (blue circle markers). Another pattern matching method using NCC to detect RNs template is then applied to reduce the drift error propagation. When NCC value is local maximum and lager then the threshold, the camera trajectory will adjust the current position to the corresponding RN. The drift errors are equally distributed to the trajectory. Green square markers in Figure 2 demonstrated the motion path adjustment with NCC matched RNs from filtered trajectory. The accumulated drift error can be zeroing when passing through each RN.

Rotation angles between adjacent sampling frames were calculated as illustrated in the left part of Figure 3 for the two cases. By accumulating the rotation angles, three corner positions (B, C and D) can be easily detected and calculated. Similar result also can be achieved by analyzing the IMU data (right part of Figure 3). As displayed in the figure, the quality of rotation angles obtained from IMU data are stable than from the video frames, which in term affects the trajectory estimation results in Figure 2.

The Velocity Random Walk (VRW) results in a massive drift error propagation using the time-integration trajectory estimated method. To address this issue, this research combined translation values calculated from the vision-based method and rotation angles received from MEMS-based IMU. Fusing the estimations of vision-based results and the IMU rotation angles registered with matched RNs provides higher accuracy of the computed trajectory as demonstrated in Figure 2 (triangle markers in magenta).

5. CONCLUDING REMARKS

This study developed a reduced-complexity vision-assisted approach and combined with filtering mechanism, low-cost MEMS-based IMU data and RNs registration for trajectory estimation and tracking. Orientation and translation parameters of the mobile device can be estimated from video frames or highly over-lapped images acquired with built-in camera of the mobile device. MEMS-based IMU can provide high-frequency and contiguous velocity and orientation information. The developed filtering mechanism can successfully remove blunders and the drift-effect caused by video vibration or instability (shaking) of the device. Registration of RNs can calibrate the trajectory and alleviate the error propagation problem. In order to improve the accuracy of trajectory estimation using MEMS-based IMU data, a data fusion scheme is proposed combining translation values from vision-based results and rotation angles from IMU measurement. Under the circumstance of no available RNs along the path, the developed fusion algorithm can mitigate the IMU VRW and vision-based rotation angle uncertainty problem. Future works will focus on integrating the displacements estimated from MEMS-based IMU to replace unreliable vision-based translations.

ACKNOWLEDGEMENTS

References

Achtelik, M., Weiss, S. and Siegwart, R., 2011. Onboard IMU and monocular vision based control for MAVs in unknown in- and outdoor environments. In: 2011 IEEE International Conference on Robotics and Automation, pp. 3056–3063.

Chang, H. and Tsai, F., 2013. Combining vision-based trajectory and wireless sensor network for location estimation and tracking. In: 2013 International Symposium on Remote Sensing, Chiba, Japan.

Ching, J., Domingo, C., Iglesia, K., Ngo, C. and Chua, N., 2013. Mobile indoor positioning using wi-fi localization and image processing. In: S.-y. Nishizaki, M. Numao, J. Caro and M. Suarez (eds), Theory and Practice of Computation, Springer, Japan, pp. 242–256.

Davison, A. J., Reid, I. D., Molton, N. D. and Stasse, O., 2007. MonoSLAM: Real-time single camera SLAM. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, pp. 1052–1067.

Fakharian, A., Gustafsson, T. and Mehrfam, M., 2011. Adaptive Kalman filtering based navigation: An IMU/GPS integration approach. In: 2011 IEEE International Conference on Networking, Sensing and Control (CNSC2011), pp. 181–185.

Gu, Y., Lo, A. and Niemegeers, I., 2009. A survey of indoor positioning systems for wireless personal networks. IEEE Communications Surveys and Tutorials 11, pp. 13–32.

Hide, C., Botterill, T. and Andreotti, M., 2010. Low cost vision-aided IMU for pedestrian navigation. In: 2010 IEEE Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), pp. 1–7.

Hightower, J. and Borriello, G., 2001. A survey and taxonomy of location systems for ubiquitous computing. IEEE Computers 34, pp. 57–66.

Iddan, G. J. and Yahav, G., 2001. 3d imaging in the studio (and elsewhere...). Proceedings of SPIE 4928, pp. 48–55.

Kelly, J., Saripalli, S. and Sukhatme, G. S., 2008. Combined visual and inertial navigation for an unmanned aerial vehicle. In: Field and Service Robotics, Springer, pp. 255–264.

Kohoutek, T. K., Droeschel, D., Mautz, R. and Behnke, S., 2013. Indoor positioning and navigation using time-of-flight cameras. In: TOF Range-Imaging Cameras, Springer, pp. 165–176.

Scaramuzza, D., Fraundorfer, F., Pollefeys, M. and Siegwart, R., 2009. Absolute scale in structure from motion from a single ve-hicle mounted camera by exploiting nonholonomic constraints. In: 12th IEEE International Conference on Computer Vision, pp. 1413–1419.

Strasdat, H., Montiel, J. and Davison, A., 2010. Scale drift-aware large scale monocular SLAM. In: Robotics: Science and Systems, Vol. VI.

Tsai, F., Chiou, Y.-S. and Chang, H., 2013. A positioning scheme combining location tracking with vision assisting for wireless sensor networks. Journal of Applied Research and Technology 11, pp. 292–300.

Woodman, O. J., 2007. An introduction to inertial navigation. Technical Report UCAMCL-TR-696, Computer Laboratory, University of Cambridge.