Autonomous Integrity Monitoring of a vision based pedestrian dead reckoning system

Yuqin Wang;Biyu Tang; Lingxiang Zheng*; Zhichao Lin
School of Information Science and Engineering
Xiamen University, Xiamen, China
lxzheng@xmu.edu.cn

Abstract. With the increasing demand for indoor positioning, VIO (visual inertial odometer) are receiving more and more attention due to the universality and convenience of the camera. We found that the visual observation of VIO is more susceptible to the environment, and the error of observation results in the final positioning error. For this reason, we specifically analyzed the visual error source for different scenarios and use the short-time reliability of PDR（Pedestrian Dead Reckoning). PDR is used to assist in the monitoring of integrity of visual observations.

Keywords: visual inertial odometer, Pedestrian Dead Reckoning, autonomous integrity monitoring.

1 Introduction

As people become more and more dependent on services of location, the need for more accurate indoor location services is becoming more and more urgent. So indoor positioning technologies based on various types of sensors (such as Wi-Fi[1], Bluetooth[2], cameras[3], inertial sensors[4], etc.) are rapidly developing. VO (visual odometer) are receiving more and more attention due to the universality and convenience of the camera. However, a visual odometer that uses a camera is difficult to handle dynamic obstacles, inertial navigation can mitigate the effects of dynamic objects by receiving motion information. At the same time, the IMU (Inertial Measurement Unit) can measure the angular velocity and acceleration that there is a significant drift in these measurements, which makes the pose obtained by the two times of integration of IMU very unreliable. The camera data can effectively estimate and correct drift caused by IMU. The VIO provides a reliable choice for indoor location services that based on the complementarity between the IMU and the camera. So camera and IMU have been used in many INS(Indoor Navigation System)[5, 6] .
According to their pose estimation methods, visual inertia odometers can be divided into two categories: the filter estimation method and the optimization-based estimation method. Filter-based classic VIOs include MSCKF (Multi-State Constraint Kalman Filter)[7] and ROVIO (Robust Visual Inertial Odometry)[8], and the VIO based on optimization method is the VINS (Visual-Inertial State Estimator)[9]. The indoor

* Corresponding author.
lxzheng@xmu.edu.cn(Lingxiang Zheng)

CEUR-WS.org/Vol-2498/short61.pdf

positioning system that based on PDR only require the data of IMU. The system uses the physiological characteristics of pedestrian walking to estimate the pedestrian's trajectory by detecting the step size and stride frequency of pedestrian walking. The source of information for the VIO system is the binocular camera and the IMU, while the PDR has only one source of information is IMU. When using the VIO system for indoor pedestrian positioning experiments, they unable to obtain accurate visual observations in special environments (such as low light and light imbalance) which resulting in large positioning errors. We analyzed the visual measurement part and positioning errors of VIO. And we proposed autonomous integrity monitoring of a vision based pedestrian dead reckoning system. Which the short-term reliability of PDR (Pedestrian Dead Reckoning)[10] is used to assist in the visual integrity detection of VIO systems.

2 Background

Visual inertial odometers are generally divided into fore-end and back-end. The fore-end mainly deals with sensor's observations, such as feature extraction and integration of images with IMU data. The back-end mainly processes the residuals caused by the observations through a filter or an optimization scheme, and obtains an optimal position state. The structural block diagram of the system is shown in Fig.1.

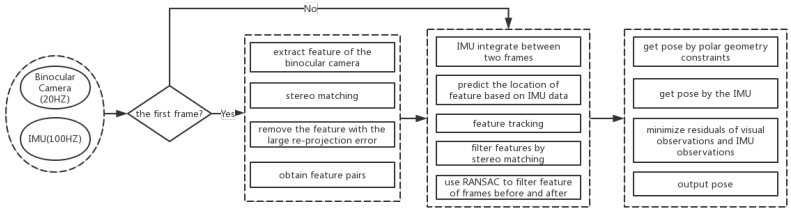

Fig. 1. The full pipeline of VIO

The k-th state vector X_k in visual inertial odometer can be defined as $\mathbf{X}_k$ including the state $\mathrm{X}_{IMU_k}^T$ of the IMU and a camera pose （attitude $\delta\theta_{C_k}^T$ and position ${}^G p_{C_k}^T$）. The measurement model of the feature point is expressed by the following equation

$$\mathbf{X}_k = \begin{bmatrix} \mathbf{X}_{IMU_k}^T & \boldsymbol{\delta\theta}_{C_k}^T & {}^{G\sim T}\mathbf{p}_{C_k} \end{bmatrix}^T, \quad \mathbf{z}_i = \frac{1}{{}^C Z_i}\begin{bmatrix} {}^C X_i \\ {}^C Y_i \end{bmatrix} + \mathbf{n}_i, i \in f, C \in \{C_1, C_2\} \tag{1}$$

where f represents a collection of all feature points, C_1, C_2 are represent the left and right cameras respectively, and $\mathbf{n}_i$ is the 2×1 image noise vector. The feature position expressed in the camera frame, ${}^C\mathbf{p}_i$,is given by:

$$ {}^{C}\mathbf{p}_i = \begin{bmatrix} {}^{C}X_i \\ {}^{C}Y_i \\ {}^{C}Z_i \end{bmatrix} = \mathbf{C}\left({}^{C_i}_{G}\bar{q}\right)\left({}^{G}\mathbf{p}_i - {}^{G}\mathbf{p}_C\right) \tag{2} $$

where ${}^{G}\mathbf{p}_i$ is the 3D feature position in the global frame, ${}^{G}\mathbf{p}_C$ is the camera in the global frame and $\mathbf{C}\left({}^{C_i}_{G}\bar{q}\right)$ is the rotation matrix between the camera frame and the global frame. Once the estimate of the feature position is obtained, we can compute the measurement residual:

$$ \mathbf{r}_i = \mathbf{z}_i - \hat{\mathbf{z}}_i \simeq \mathbf{H}_C\mathbf{X} + \mathbf{H}_i\,{}^{G}\tilde{\mathbf{p}}_i + \mathbf{n}_i, \quad where \begin{cases} \mathbf{H}_C = \dfrac{\partial \mathbf{z}_i}{\partial {}^{C_1}\mathbf{p}_i}\cdot\dfrac{\partial {}^{C_1}\mathbf{p}_i}{\partial \mathbf{X}_{C_1}} + \dfrac{\partial \mathbf{z}_i}{\partial {}^{C_2}\mathbf{p}_i}\cdot\dfrac{\partial {}^{C_2}\mathbf{p}_i}{\partial \mathbf{X}_{C_1}} \\ \mathbf{H}_i = \dfrac{\partial \mathbf{z}_i}{\partial {}^{C_1}\mathbf{p}_i}\cdot\dfrac{\partial {}^{C_1}\mathbf{p}_i}{\partial {}^{G}\mathbf{p}_i} + \dfrac{\partial \mathbf{z}_i}{\partial {}^{C_2}\mathbf{p}_i}\cdot\dfrac{\partial {}^{C_2}\mathbf{p}_i}{\partial {}^{G}\mathbf{p}_i} \end{cases} \tag{3} $$

where $\mathbf{H}_C$ and $\mathbf{H}_i$ are the Jacobians of the measurement $\mathbf{z}_i$ with respect to the state and the position estimate of feature. With all the sets of measurement equations formed by the feature points, we can get the optimal solution by minimizing the error and get the optimal position estimate.

3 Visual error analysis & Autonomous integrity monitor

3.1 Visual error analysis

We found that the system has a large error in the positioning results obtained under special environments during the experiment. So we analyzed the error source and divided it into the following four error situations.

Insufficient features. Commonly used feature extraction algorithms include SIFT[11], SUFT[12], FAST[13], ORB[14] and other feature extraction algorithms. And those feature extraction algorithm is often used in processing of VIO projects. In fact, it can be derived from the definition that it is desirable to find a point with strong contrast with surrounding pixels as a feature point. The contrast of point P can be expressed as

$$ \mathrm{V}(x,y) = \sum\left|I(x+\triangle x, y+\triangle y) - I(x,y)\right|, \quad \mathrm{V}(x,y) \sim \sum\left[I_x{}^2\cdot\triangle x^2 + I_y{}^2\cdot\triangle y^2 + 2I_xI_y\cdot\triangle x\triangle y\right] \tag{4} $$

The value of V mainly depends on the gradient value of the point P in the x and y directions. The larger the gradient value, the easier it is to be detected by the detector. Scenes with less texture (white wall) and low light in the indoor environment are very common, so that the detector could not detect enough feature points. Position estimation can be performed when the feature point pairs exceed 9 pairs.

$$ \begin{bmatrix} \delta\hat{\mathbf{X}} \\ \delta\,{}^{G}\hat{\mathbf{p}}_i \end{bmatrix} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{r} \rightarrow \hat{\mathbf{X}} = \mathbf{X} + \delta\hat{\mathbf{X}}, \quad {}^{G}\hat{\mathbf{p}}_i = {}^{G}\mathbf{p}_i + \delta\,{}^{G}\hat{\mathbf{p}}_i \tag{5} $$

When the number of feature points is sufficient, $\text{rank}(\mathbf{H}) \geq 9$,the constraint equation is sufficient to obtain the optimal solution. When the number of feature points is insufficient, the constraint condition is insufficient, and the obtained **X** error is large. This leads to an increase in positioning error.

Lighting causes the failure of feature tracking. Light changes often occur in indoor environments and we use the Lambertian model as the lighting model.

$$I(x,y) = \rho(x,y) \cdot h(x,y)^T \cdot S \tag{6}$$

Where $\mathrm{I}(x,y)$ is the image gray value, $\rho(x,y)$ is the object reflectivity, $h(x,y)$ is the surface normal vector and S is the light intensity. We found that feature tracking is easy to lose leads to inaccurate positioning during light changing. The assumption of the optical flow method assumes that the gray level is unchanged. Substitute the lighting formula:

$$\frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} = -\frac{\partial I}{\partial t} \rightarrow \begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} \mu \\ \nu \end{bmatrix} = -\rho \cdot \left[\frac{\partial h^T}{\partial t} \cdot S + \frac{\partial S}{\partial t} \cdot h^T \right] \tag{7}$$

Where I_x and I_y are the gradient values of the feature points in the x and y directions, respectively, μ and ν are the velocity of the motion in the x and y directions representing the feature points. When the light intensity changes, the residual of the feature points caused by the illumination change can be expressed as

$$\delta \mathbf{r}_i = \frac{1}{{}^C Z_i} \begin{bmatrix} {}^C X_i - ({}^C X_i + \mu t) \\ {}^C Y_i - ({}^C Y_i + \nu t) \end{bmatrix} = \frac{t}{{}^C Z_i} \begin{bmatrix} I_x \\ I_y \end{bmatrix} \cdot \left[\rho \frac{\partial S}{\partial t} \cdot h^T \right] \tag{8}$$

Uneven distribution of features. It can be seen in the observation equation of the image that the presence of noise causes positional errors in the feature points in the image. The position error of the feature points will affect the state estimation of the camera when calculating the re-projection error. We describe the quality of the position estimate based on the camera state's Jacobian matrix of feature points $\mathbf{H}_i$. In our model, the measurement error is zero-average, so the positioning error is also zero-average, so we can obtain the expected value and covariance of the error in the position calculation.

$$E(\Delta \mathbf{X}) = E\left(\hat{\mathbf{X}} - \mathbf{X}\right) = 0, \quad Cov[\Delta \mathbf{X}] = \sigma^2 \left(\mathbf{H}_i{}^T \mathbf{H}_i\right)^{-1} \sqrt{b^2 - 4ac} \tag{9}$$

The amount of change in the position error in the x, y and z directions is represented by $\sigma_x{}^2, \sigma_y{}^2, \sigma_z{}^2$ respectively. Use H_{ii} to represent the first element on the diagonal in the diagonal matrix $\left(\mathbf{H}_i{}^T \mathbf{H}_i\right)^{-1}$. Then can be expressed as

$$SD(\Delta \mathbf{X}) = \sqrt{\sigma_x{}^2 + \sigma_y{}^2 + \sigma_z{}^2} = \sqrt{H_{11} + H_{22} + H_{33}} \tag{10}$$

Moving feature point. All moving objects such as pedestrians or vehicles will affect the positioning result when we positioning. When the feature points of the camera are concentrated on the moving object and the relative movement of the feature points is relatively large, the calculated camera movement is also too large. That is to say, the world coordinates of the feature points have an additional motion shift, which affects the camera's observation. Let's analyze the residuals generated by the offset of the feature.

$$\begin{bmatrix} {}^{C_i}X_j+\triangle x \\ {}^{C_i}Y_j+\triangle y \\ {}^{C_i}Z_j+\triangle z \end{bmatrix} = \mathbf{C}\left({}^{C_i}_{G}\bar{q}\right)\left({}^{G}\mathbf{p}_{f_j}+\triangle {}^{G}\mathbf{p}_{f_j}-{}^{G}\mathbf{p}_{C_i}\right)\rightarrow \quad \mathbf{r}_i=\frac{\triangle z}{{}^{C}Z_i({}^{C}Z_i+\triangle z)}\begin{bmatrix} {}^{C}X_i \\ {}^{C}Y_i \end{bmatrix}-\frac{1}{{}^{C}Z_i+\triangle z}\begin{bmatrix} \triangle x \\ \triangle y \end{bmatrix} \quad (11)$$

Autonomous integrity monitoring

The error bound of PDR. Although PDR has a problem of cumulative error, the error in a short time is very small. Now assume that there are two sampling points O_1, O_2 , the sampling interval is Δt and the velocity of time O_1 is v , the displacement is s , and the state covariance matrix is $P_1=\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$.Accelerometer observations can cause inaccurate deviations due to shocks generated during motion. We first analyze the one-dimensional motion, the acceleration at time O_1 is $f_{mea}=f_{true}+\delta f$ .An estimated value of the state quantity at O_2 can be obtained from the state quantity at O_1 . The deviation caused by δf is:

$$\begin{bmatrix} \delta s \\ \delta v \end{bmatrix}=\begin{bmatrix} \delta f\cdot\Delta t\left(\frac{1}{2}\Delta t-\frac{p_{12}+\Delta t\cdot p_{22}}{p_{22}+R}\right) \\ \delta f\cdot\Delta t\left(1-\frac{p_{22}}{p_{22}+R}\right) \end{bmatrix} \quad (12)$$

Where R is the covariance matrix of the observed noise.

PDR-assisted visual integrity monitoring. Hypothesis deviation obeys Gaussian distribution $\mathbf{e}\sim N(0,\Sigma)$. Now $\mathbf{e}$ is a three-dimensional vector. In order to facilitate the calculation, the inner product of the computation vector is transformed into a scalar .

$$r=\mathbf{e}^T\Sigma\mathbf{e}=\left(\Sigma^{-\frac{1}{2}}\mathbf{e}\right)^T\left(\Sigma^{-\frac{1}{2}}\mathbf{e}\right), where\left(\Sigma^{-\frac{1}{2}}\mathbf{e}\right)\sim N(\mathbf{0},\mathbf{I}) \quad (13)$$

It can be thought of as the sum of the squares of two independent random variables subject to the standard normal distribution, which obeys the chi-square distribution of three degrees of freedom. The probability distribution (cumulative distribution function) is $\alpha=F(x)$, and given a α , we can determine an interval $\left[0,F^{-1}(\alpha)\right]$, and $F^{-1}(\alpha)$ is the threshold we are looking for to determine the visual integrity.

4 Experimental & Evaluation

Insufficient features. As shown in Fig.2(a), We changed the feature point extraction threshold for the same set of data, and the control variable is the number of features per frame. Moreover, the number of feature points is 0 at the 1560-th frame, because the white wall is encountered and the feature points cannot be extracted. We draw the corresponding positioning trajectory. As shown in Fig.2(b). When the feature points are scarce, the camera's ability to correct the IMU is not strong, the path is not serrated enough, and the trajectory also shows significant deviations between the x-axis and the y-axis.

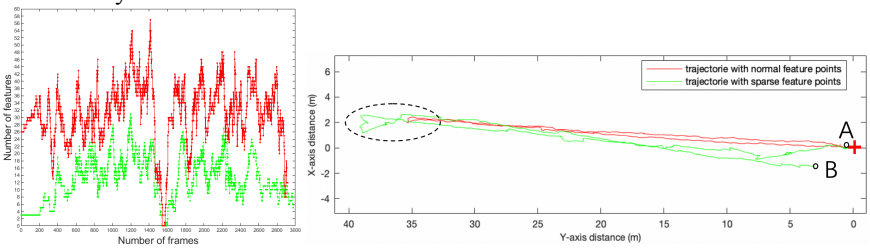

Fig. 2. (a)Red is the number of features with the threshold is 20 per frame, and green is the number of sparse features with the threshold is 60 per frame. (b) The red cross mark is the starting point, the point A is the end point of the track in the original state of the feature point, and the point B is the end point of the track in the sparse state of the feature point.

Lighting causes the failure of feature tracking. When the light is different from the left and right cameras, the average gray value of the images acquired by the left and right cameras is different, and the matching rate is low. The Fig.3(a) is a feature point distribution map obtained by FAST feature extraction on the images acquired by the left and right cameras. However, the image matching rate of the left and right cameras is not high, and the matching ratio is only 0.55. No feature points exist in the image after stereo matching. If the feature detection module does not have the feature point data output, the visual inertia mileage calculation method cannot perform the posture update, resulting in the track accumulating offset, and serious errors may occur. As shown in Fig.3(b).

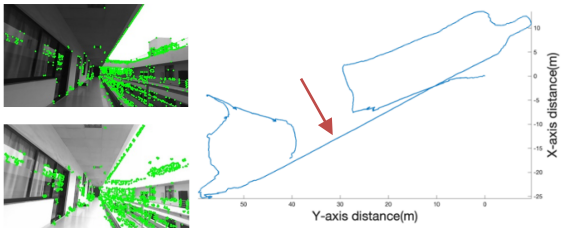

Fig. 3. (a) The average gray value of the top image is 97.0845, the average gray value of the bottom image is 183.946, and the number of features extracted by the top image is 946. The number of features extracted by the bottom image is 1543. (b) A long line pointed by the red arrow is that the different light of the left and right cameras results in no feature points and serious deviation of the trajectory

Uneven distribution of features. We use the distribution of feature points as variables and compare the trajectory with the original feature distribution. As shown in Fig.4(a). However, when the feature points are only distributed in the red area, the movement trajectory of the feature points is directed to the right side of the image. As shown in Fig.4(b), it can be seen that the trajectory of the feature with uneven distribution has obvious deviation to the left.

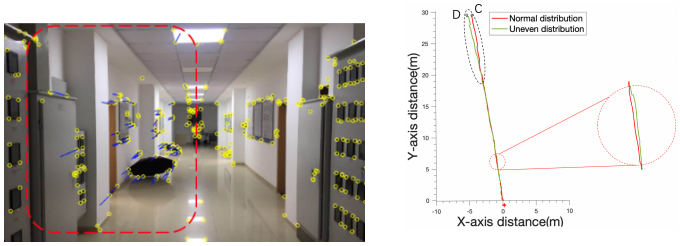

Fig. 4. (a) The feature point distribution is controlled in the red area. The yellow circle represents all the extracted feature points, and the blue line segment represents the tracking track of the feature points. (b) The red cross mark is the starting point, the C point is the end point of the track where the feature point distribution is normal, and D is the track end point where the feature point is unevenly distributed.

Moving feature point. The pedestrians pass in front of the camera, and the contrast track is drawn as shown in Fig.5(b). It is obvious in the circle that the green track is shifted to the left because of the influence of pedestrians. We analyzed the details of this moment. As can be seen from Fig.5(a), when the pedestrian moves, more than half of the extracted feature points are gathered on the pedestrian. Therefore, the movement of pedestrian relative to the camera will lead to the deviation of the positioning results.

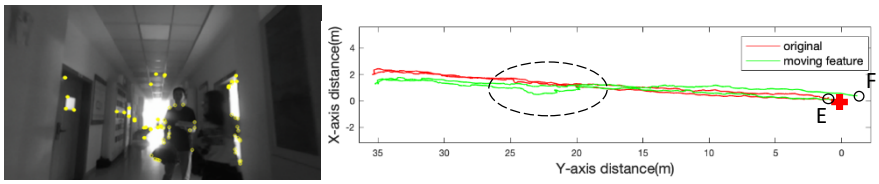

Fig. 5. (a) Open circles represent moving feature points, solid circles represent stationary feature points. (b)The red cross is marked as the starting point and the E point is the original track. End point, F is the end point of the trajectory affected by the moving feature points.

5 Conclusion

In this paper, we analyzed the influences of visual measurements based on the positioning error of the system of visual inertial odometer, and analyzed the source of its positioning error. We divided the error sources of visual measurement into four cases and we have confirmed the influence of four error sources by experiments. Through the error analysis of PDR, it is found that the error of PDR in a short time is small and bounded. According to that characteristics of PDR, it is used to assist in the detection

of visual errors. When the above four cases of visual errors are detected, the positioning result will be corrected.

References

1. S. He and S. G. Chan, "Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons," IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 466-490, 2016.
2. G. D. Blasio, A. Quesada-Arencibia, C. R. García, J. C. Rodríguez-Rodríguez, and R. Moreno-Díaz, "A Protocol-Channel-Based Indoor Positioning Performance Study for Bluetooth Low Energy," IEEE Access, vol. 6, pp. 33440-33450, 2018.
3. Z. Feng and S. Hao, "Low-Light Image Enhancement by Refining Illumination Map with Self-Guided Filtering," in 2017 IEEE International Conference on Big Knowledge (ICBK), 2017, pp. 183-187.
4. W. Elloumi, A. Latoui, R. Canals, A. Chetouani, and S. Treuillet, "Indoor Pedestrian Localization With a Smartphone: A Comparison of Inertial and Vision-Based Methods," IEEE Sensors Journal, vol. 16, no. 13, pp. 5376-5388, 2016.
5. L. Mainetti, L. Patrono, and I. Sergi, "A survey on indoor positioning systems," in 2014 22nd International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2014, pp. 111-120.
6. S. Garcia-Villalonga and A. Perez-Navarro, "Influence of human absorption of Wi-Fi signal in indoor positioning with Wi-Fi fingerprinting," in 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2015, pp. 1-10.
7. A. I. Mourikis and S. I. Roumeliotis, "A Multi-State Constraint Kalman Filter for Vision-Aided Inertial Navigation," 2007.
8. M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct EKF-based approach," in IEEE/RSJ International Conference on Intelligent Robots & Systems, 2015.
9. Q. Tong, P. Li, and S. J. I. T. o. R. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," vol. PP, no. 99, pp. 1-17, 2017.
10. D. Kamisaka, S. Muramatsu, T. Iwamoto, H. J. I. T. o. I. Yokoyama, and Systems, "Design and Implementation of Pedestrian Dead Reckoning System on a Mobile Phone," vol. 94-D, no. 6, pp. 1137-1146, 2011.
11. D. G. J. I. J. o. C. V. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," journal article vol. 60, no. 2, pp. 91-110, November 01 2004.
12. H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded Up Robust Features," in European Conference on Computer Vision, 2006.
13. C. Henderson and E. Izquierdo, "Robust Feature Matching in Long-Running Poor-Quality Videos," IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 6, pp. 1161-1174, 2016.
14. E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski, "ORB: an efficient alternative to SIFT or SURF," in International Conference on Computer Vision, 2012.