THE JOURNAL OF NAVIGATION (2018), **71**, 1329–1343. © The Royal Institute of Navigation 2018
doi:10.1017/S0373463318000310

Minimum Sigma Set SR-UKF for Quadrifocal Tensor-based Binocular Stereo Vision-IMU Tightly-coupled System

Maosong Wang[1,2], Wenqi Wu[1], Naser El-Sheimy[2] and Zhiwen Xian[3]

[1](*National University of Defense Technology, China*)
[2](*The University of Calgary, Canada*)
[3](*Taiyuan Satellite Launch Center, China*)
(E-mail: wangmaosong12@hotmail.com)

This paper presents a binocular vision-IMU (Inertial Measurement Unit) tightly-coupled structure based on a Minimum sigma set Square-Root Unscented Kalman Filter (M-SRUKF) for real time navigation applications. Though the M-SRUKF has only half the sigma points of the SRUKF, it has the same accuracy as the SRUKF when applied to the binocular vision-IMU tightly-coupled system. As the Kalman filter flow is a kind of square-root system, the stability of the system can be guaranteed. The measurement model and the outlier rejection model of this tightly-coupled system not only utilises the epipolar constraint and the trifocal tensor geometry constraint between the consecutive two image pairs, but also uses the quadrifocal tensor geometry among four views. The structure of the binocular vision-IMU tightly-coupled system is in the form of an error state, and the time updates of the state and the state covariance are directly estimated without using Unscented Transformation (UT). Experiments are carried out based on an outdoor land vehicle open source dataset and an indoor Micro Aerial Vehicle (MAV) open source dataset. Results clearly show the effectiveness of the proposed new mechanisation.

KEYWORDS

1. Minimum Sigma Set Square-Root Unscented Kalman Filter. 2. Multiple View Geometry. 3. Quadrifocal Tensor. 4. Binocular Tightly-coupled Vision-IMU System.

Submitted: 8 October 2017. Accepted: 22 April 2018. First published online: 13 June 2018.

1. INTRODUCTION. Vision-aided inertial navigation systems have been extensively studied (Kelly and Sukhatme, 2011; Feng et al., 2012; Weiss et al., 2012; Kottas and Roumeliotis, 2013; Hesch et al., 2014; Xian et al., 2015; Liu et al., 2018), as IMUs and vision systems have complementary features. For example, IMUs have the ability to capture fast ego-motion with lower uncertainty, while cameras can track feature points precisely in slow motion. This kind of integrated system can operate in environments where

https://doi.org/10.1017/S0373463318000310 Published online by Cambridge University Press

Global Navigation Satellite Systems (GNSS) are unreliable or unavailable. Corke et al. (2007), made a comprehensive introduction of loosely-coupled and tightly-coupled vision and IMU integration systems. Loosely-coupled vision and IMU integration systems fuse the inertial navigation solution with relative pose estimation between consecutive image measurements (Tardif et al., 2010; Sirtkaya et al., 2013). Tightly-coupled methods, in contrast, process the raw information of both sensors in a single estimator, thus leading to higher accuracy (Mourikis and Roumeliotis, 2007). A kind of vision-aided inertial tightly-coupled mechanisation which is based on tri-focal tensor geometry has been studied by many researchers (Asadi and Bottasso, 2012; Indelman et al., 2012; Fang and Huang, 2013). For feature extraction, it is better to use both the lines and the points, because they can provide complementary information about the image (Zhang, 2013; Kitt et al., 2010). Hu and Chen (2014) proposed a sliding window odometry using monocular camera geometry constraints among three images as measurements; the results showed that the mechanisation worked well, when point and line features were used. Experiments by Kong et al. (2015), also showed that by using both point and line features, navigation accuracy was better than by merely using point features. An Unscented Kalman Filter (UKF) was used in the works by Hu and Chen (2014) and Kong et al. (2015).

However, UKF may have some problems, since it cannot guarantee the stability of the tightly-coupled system, and it has to re-determine the positive definiteness of the estimation covariance matrix. Square-Root Unscented Kalman Filter SRUKF (Merwe et al., 2001) is an improvement of UKF, which uses QR decomposition (where Q is an orthogonal matrix and R is an upper triangular matrix) instead of Cholesky decomposition as in UKF. This can not only avoid the most expensive operation by merely calculating the square-root of the state covariance at each time updating, but also guarantees the positive semi-definiteness of the state covariance to ensure the stability of the computation. From this perspective, SRUKF is a better filter method to replace UKF. Based on this fact, Xian et al. (2016), presented a SRUKF-based binocular stereo Vision/IMU tightly-coupled system, among which multiple view geometry constraints were used. The results indicated that this mechanisation provided benefits over that identified in Hu and Chen (2014).

However, the standard SRUKF uses $2n + 1$ sigma points for the UT which will affect the computation time. UT methods have some variants with reduced sigma sets. For example, the simplex sigma set uses $n + 2$ sigma points (Julier et al., 2002). However, this kind of UT method has the problem that the radius which bounds the sphere of the sigma points is $2^{n/2}$. Therefore, at even relatively low dimensions, there are potential problems with numerical stability. The spherical simplex sigma set uses $n + 2$ sigma points (Julier et al., 2003). The sigma points lie in the hypersphere with a radius of $\sqrt{n/(1 - W_0)}$, and this had better stability than the earlier proposal. Menegaz et al. (2011), proposed a minimum sigma set UT by using only $n + 1$ sigma points, and the experiments in this paper showed that this algorithm worked well. A novel simplex unscented transform and filter was proposed by Li et al. (2007), which also used $n + 1$ sigma points. The UKF and SRUKF family were systematically illustrated by Menegaz et al. (2015). There are some conceptual statements in the reduced sigma point set section, and they are revised through theoretical analysis (Wang et al., 2017).

This paper introduces the quadrifocal tensor constraints in a binocular Vision-IMU tightly-coupled system. As far as the authors are aware, this is the first time that a quadrifocal tensor has been used in a real time positioning and navigation system. The structure of the binocular vision-IMU tightly-coupled system is in the form of error state, and the time

https://doi.org/10.1017/S0373463318000310 Published online by Cambridge University Press

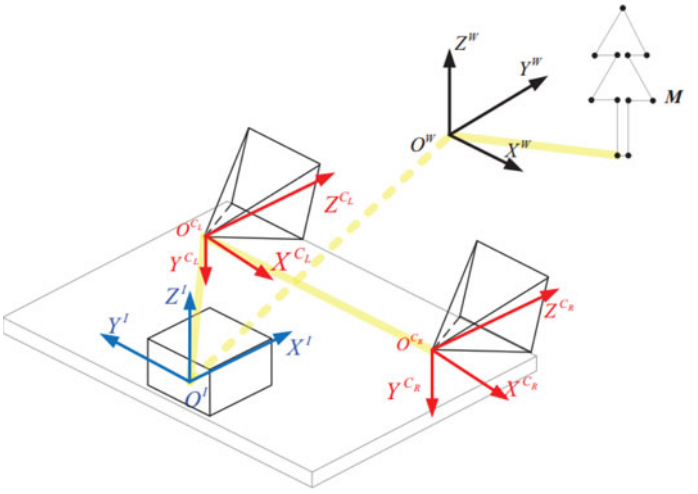


Figure 1. Reference frames and their relationships.

updates of the state and the state covariance are directly estimated without using UT transformation. In order to further save computation cost, M-SRUKF is used in the binocular vision-IMU tightly-coupled system.

Several ego-motion cases from both outdoor and indoor open source datasets are used for evaluating the performance of the new algorithms. Results showed that the new M-SRUKF based binocular vision-IMU mechanisation by using four views geometry performs well in all the test cases.

2. VISION-IMU TIGHTLY-COUPLED SYSTEM MODEL.

The reference frames and system model used in the vision-IMU tightly-coupled system are defined as follows (Xian et al., 2016).

Figure 1 shows the relationships of the three reference frames. The W frame is the world frame, where the pose of the IMU is estimated with respect to this frame. The C frame is the camera frame, where C_L is the left camera, C_R is the right camera, and they are rigidly attached. The I frame is the IMU frame, with origin at the centre of the IMU body. The feature points M are modelled in the world frame W.

The IMU state vector is defined as:

$$\mathbf{x}_{IMU}(t) = \left[\left(\bar{\boldsymbol{q}}_{WI}\right)^T \quad \left(\boldsymbol{v}_{WI}^W\right)^T \quad \left(\boldsymbol{p}_{WI}^W\right)^T \quad \left(\boldsymbol{b}_g(t)\right)^T \quad \left(\boldsymbol{b}_a(t)\right)^T \quad {}^1\left(\bar{\boldsymbol{q}}_{WI}\right)^T \quad {}^1\left(\boldsymbol{p}_{WI}^W\right)^T \right]^T \tag{1}$$

where $\boldsymbol{p}_{WI}^W(t)$ represents the position of IMU in the world frame $\{W\}$; $\bar{\boldsymbol{q}}_{WI}(t)$ is the unit quaternion of the IMU frame $\{I\}$ in the world frame; $\boldsymbol{v}_{WI}^W$ is the linear velocity of the IMU in the world frame; $\boldsymbol{b}_g(t)$ and $\boldsymbol{b}_a(t)$ are the IMU gyroscope and accelerometer biases, respectively. $\left({}^1\left(\bar{\boldsymbol{q}}_{WI}\right)^T, {}^1\left(\boldsymbol{p}_{WI}^W\right)^T\right)$ denotes the last pose of the IMU respect to the frame $\{W\}$.

The biases $\boldsymbol{b}_g(t)$ and $\boldsymbol{b}_a(t)$ are modelled as a Gaussian random walk process, driven by the white, zero-mean noise vectors $\boldsymbol{n}_{gw}$ and $\boldsymbol{n}_{aw}$, with covariance matrices $\boldsymbol{Q}_{gw}$ and $\boldsymbol{Q}_{aw}$, respectively. The time evolution of the IMU state is given by the following equation (Kelly

https://doi.org/10.1017/S0373463318000310 Published online by Cambridge University Press

et al., 2011):

$$\dot{\boldsymbol{p}}_{WI}^{W} = \boldsymbol{v}_{WI}^{W}, \dot{\bar{\boldsymbol{q}}}_{WI} = \frac{1}{2}\Omega\left(\boldsymbol{\omega}_{WI}^{I}\right)\bar{\boldsymbol{q}}_{WI}, \quad \dot{\boldsymbol{v}}_{WI}^{W} = \boldsymbol{a}_{WI}^{W}, \dot{\boldsymbol{b}}_{g} = n_{gw}, \dot{\boldsymbol{b}}_{a} = \boldsymbol{n}_{aw} \tag{2}$$

where the $\Omega\left(\boldsymbol{\omega}_{WI}^{I}\right)$ is the quaternion multiplication matrix shown in Equation (3):

$$\Omega\left(\boldsymbol{\omega}_{WI}^{I}\right) = \begin{bmatrix} 0 & -\left(\boldsymbol{\omega}_{WI}^{I}\right)^{T} \\ \boldsymbol{\omega}_{WI}^{I} & -\left[\boldsymbol{\omega}_{WI}^{I}\times\right] \end{bmatrix} \tag{3}$$

where $\boldsymbol{\omega}_{WI}^{I}$ is the angular velocity of the IMU with respect to the world frame, and $\boldsymbol{a}_{WI}^{W}$ is the acceleration of the IMU with respect to the world frame expressed in the world frame. The measured angular velocity and linear acceleration are:

$$\boldsymbol{\omega}_{m} = \boldsymbol{\omega}_{WI}^{I} + \boldsymbol{b}_{g} + \boldsymbol{n}_{g} \tag{4}$$

$$\boldsymbol{a}_{m} = \boldsymbol{R}^{T}\left(\bar{\boldsymbol{q}}_{WI}\right)\left(\boldsymbol{a}_{WI}^{W} - \boldsymbol{g}^{W}\right) + \boldsymbol{b}_{a} + \boldsymbol{n}_{a} \tag{5}$$

where $\boldsymbol{R}\left(\bar{\boldsymbol{q}}_{WI}\right)$ is the direction cosine matrix corresponding to the unit quaternion $\bar{\boldsymbol{q}}_{WI}$, $\boldsymbol{n}_{g}$ and $\boldsymbol{n}_{a}$ are measurements of the noise of the gyroscope and accelerometer, which are assumed to be zero-mean Gaussian noise with covariance matrices $\boldsymbol{Q}_{g}$ and $\boldsymbol{Q}_{a}$, respectively. In the models, the Earth's rotation rate was not considered in the gyroscope measurement, because it is small enough relative to the noise and bias of the low-cost gyroscope.

The filter error state vector for the Kalman filter is defined as:

$$\delta\mathbf{x} = \left[\left(\delta\boldsymbol{\theta}^{I}\right)^{T} \quad \left(\delta\boldsymbol{v}_{WI}^{W}\right)^{T} \quad \left(\delta\boldsymbol{p}_{WI}^{W}\right)^{T} \quad \delta\boldsymbol{b}_{g}^{T} \quad \delta\boldsymbol{b}_{a}^{T} \quad {}^{1}\left(\delta\boldsymbol{\theta}^{I}\right)^{T} \quad {}^{1}\left(\delta\boldsymbol{p}_{WI}^{W}\right)^{T}\right]^{T} \tag{6}$$

The differential equation for the previous pose $\left({}^{1}\left(\bar{\boldsymbol{q}}_{WI}\right)^{T}, {}^{1}\left(\boldsymbol{p}_{WI}^{W}\right)^{T}\right)$ is defined as indicated in Equation (7):

$${}^{1}\dot{\boldsymbol{q}}_{WI} = 0, {}^{1}\dot{\boldsymbol{p}}_{WI}^{W} = 0 \tag{7}$$

The error-state process model is defined as:

$$\delta\dot{\mathbf{x}} = F\delta\mathbf{x} + \boldsymbol{n} \tag{8}$$

where:

$$\begin{aligned} F &= \begin{pmatrix} F_{11} & \mathbf{0}_{15\times6} \\ F_{21} & \mathbf{0}_{6\times6} \end{pmatrix} \\ \boldsymbol{n} &= \left(\mathbf{0}_{9\times1}^{T} \quad \boldsymbol{n}_{g}^{T} \quad \boldsymbol{n}_{a}^{T} \quad \mathbf{0}_{6\times1}^{T}\right)^{T} \\ F_{11} &= \begin{pmatrix} \left[\left(\boldsymbol{b}_{g} - \boldsymbol{\omega}_{m}\right)\times\right] & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & -\boldsymbol{I}_{3} & \mathbf{0}_{3\times3} \\ \boldsymbol{R}\left(\bar{\boldsymbol{q}}_{WI}\right)\left[\left(\boldsymbol{b}_{a} - \boldsymbol{a}_{m}\right)\times\right] & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & -\boldsymbol{R}\left(\bar{\boldsymbol{q}}_{WI}\right) \\ \mathbf{0}_{3\times3} & \boldsymbol{I}_{3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{6\times3} & \mathbf{0}_{6\times3} & \mathbf{0}_{6\times3} & \mathbf{0}_{6\times3} & \mathbf{0}_{6\times3} \end{pmatrix} \end{aligned} \tag{9}$$

https://doi.org/10.1017/S0373463318000310 Published online by Cambridge University Press

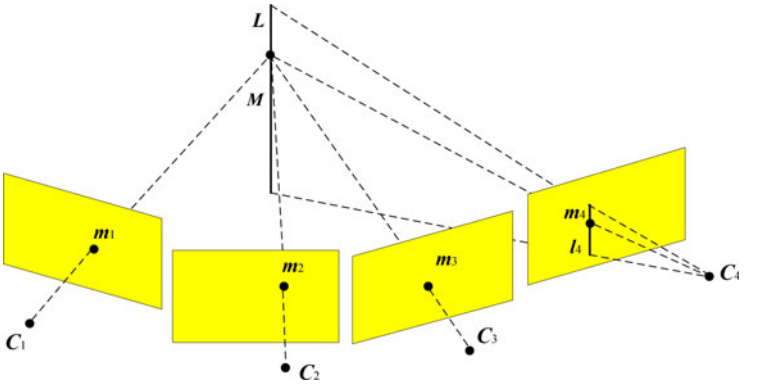


Figure 2. Point-point-point-line correspondence among four views.

$$F_{21} = \begin{pmatrix} \boldsymbol{I}_3 & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{I}_3 & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \end{pmatrix}$$

3. QUADRIFOCAL TENSOR BASED CONSTRAINTS. The measurement constraints used in this paper are the epipolar constraint between two images, the trifocal tensor constraint among three images and the quadrifocal tensor constraint among four images. The epipolar constraint and the trifocal tensor constraint are the same as described in Hu et al. (2014), Kong et al. (2015) and Xian et al. (2016). The left and right images at the earlier time are denoted respectively by Im1 and Im2, and those in the current view are denoted by Im3 and Im4. The epipolar constraints used for measurement are the pairs (Im1, Im3) and (Im2, Im4). These two pairs are chosen because they contain the relative pose information between the previous and current poses, which is helpful in improving the accuracy of the state estimates. Similarly, the trifocal tensor constraints used in the measurement are groups (Im1, Im2, Im3) and (Im1, Im2, Im4). The quadrifocal tensor constraint is among the four views (Im1, Im2, Im3, Im4).

The quadrifocal tensor geometry is shown in Figure 2. The camera matrices are defined as $\boldsymbol{P}_1 = [\boldsymbol{I}|\boldsymbol{0}]$, $\boldsymbol{P}_2 = [\boldsymbol{A}|\boldsymbol{a}_4]$, $\boldsymbol{P}_3 = [\boldsymbol{B}|\boldsymbol{b}_4]$ and $\boldsymbol{P}_4 = [\boldsymbol{C}|\boldsymbol{c}_4]$, which correspond to the different viewpoints $\boldsymbol{C}_1$, $\boldsymbol{C}_2$, $\boldsymbol{C}_3$ and $\boldsymbol{C}_4$. The normalised feature points $\tilde{\boldsymbol{m}}_1 \leftrightarrow \tilde{\boldsymbol{m}}_2 \leftrightarrow \tilde{\boldsymbol{m}}_3 \leftrightarrow \tilde{\boldsymbol{m}}_4$ across four views are considered (the feature points are normalised corresponding to $\boldsymbol{m}_1 \leftrightarrow \boldsymbol{m}_2 \leftrightarrow \boldsymbol{m}_3 \leftrightarrow \boldsymbol{m}_4$).

A common method for deriving the quadrifocal tensor is to define the linear system as (Hartley et al., 2003):

$$\begin{bmatrix} \boldsymbol{P}_1 & \tilde{\boldsymbol{m}}_1 & & & \\ \boldsymbol{P}_2 & & \tilde{\boldsymbol{m}}_2 & & \\ \boldsymbol{P}_3 & & & \tilde{\boldsymbol{m}}_3 & \\ \boldsymbol{P}_4 & & & & \tilde{\boldsymbol{m}}_4 \end{bmatrix} \begin{pmatrix} \boldsymbol{X} \\ -k_1 \\ -k_2 \\ -k_3 \\ -k_4 \end{pmatrix} = 0 \tag{10}$$

where $\boldsymbol{X}$ represents a point in Euclidean space. Since this equation has a solution, the matrix on the left in Equation (10) has rank at most 7, and so all 8×8 determinants are zero. Such a determinant leads to a quadrilinear relationship of the form:

https://doi.org/10.1017/S0373463318000310 Published online by Cambridge University Press

Table 1. Quadrilinear relations.

Correspondence	Relation
Four points	$\tilde{\boldsymbol{m}}_{1i}\tilde{\boldsymbol{m}}_{2j}\,\tilde{\boldsymbol{m}}_{3k}\tilde{\boldsymbol{m}}_{4l}\varepsilon_{ipw}\varepsilon_{jqx}\varepsilon_{kry}\varepsilon_{lsz}\mathcal{Q}^{pqrs}=0_{wxyz}$
Three points, one line	$\tilde{\boldsymbol{m}}_{1i}\tilde{\boldsymbol{m}}_{2j}\,\tilde{\boldsymbol{m}}_{3k}l_{4s}\varepsilon_{ipw}\varepsilon_{jqx}\varepsilon_{kry}\mathcal{Q}^{pqrs}=0_{wxy}$
Two points, two lines	$\tilde{\boldsymbol{m}}_{1i}\tilde{\boldsymbol{m}}_{2j}\,l_{3r}l_{4s}\varepsilon_{ipw}\varepsilon_{jqx}\mathcal{Q}^{pqrs}=0_{wx}$

where each choice of the free variables w, x, y and z gives a different equation, and the four-dimensional quadrifocal tensor $\mathcal{Q}^{pqrs}$ is defined by:

$$\mathcal{Q}^{pqrs}=\det\begin{bmatrix}\boldsymbol{p}_1^p\\ \boldsymbol{p}_2^q\\ \boldsymbol{p}_3^r\\ \boldsymbol{p}_4^s\end{bmatrix} \tag{11}$$

In this paper, point-point-point-line correspondence is used for creating the four view measurement equations. The reason for this choice is that by merely using point features, the badly distributed feature points in some frames will lead to a bad orientation estimation (Kong et al., 2015). The measurement model is created by choosing four combinations of w, x, y and z:

$$z_1=\begin{bmatrix}0_{111}\\ 0_{222}\\ 0_{333}\\ 0_{123}\end{bmatrix}=h(\boldsymbol{x}_k,\boldsymbol{m}_1,\boldsymbol{m}_2,\boldsymbol{m}_3,\boldsymbol{m}_4)=\begin{bmatrix}\tilde{\boldsymbol{m}}_{1i}\tilde{\boldsymbol{m}}_{2j}\,\tilde{\boldsymbol{m}}_{3k}l_{4s}\varepsilon_{ip1}\varepsilon_{jq1}\varepsilon_{kr1}\mathcal{Q}^{pqrs}\\ \tilde{\boldsymbol{m}}_{1i}\tilde{\boldsymbol{m}}_{2j}\,\tilde{\boldsymbol{m}}_{3k}l_{4s}\varepsilon_{ip2}\varepsilon_{jq2}\varepsilon_{kr2}\mathcal{Q}^{pqrs}\\ \tilde{\boldsymbol{m}}_{1i}\tilde{\boldsymbol{m}}_{2j}\,\tilde{\boldsymbol{m}}_{3k}l_{4s}\varepsilon_{ip3}\varepsilon_{jq3}\varepsilon_{kr3}\mathcal{Q}^{pqrs}\\ \tilde{\boldsymbol{m}}_{1i}\tilde{\boldsymbol{m}}_{2j}\,\tilde{\boldsymbol{m}}_{3k}l_{4s}\varepsilon_{ip1}\varepsilon_{jq2}\varepsilon_{kr3}\mathcal{Q}^{pqrs}\end{bmatrix} \tag{12}$$

The epipolar and the trifocal tensor constraints are defined in Equation (13) (Xian et al., 2016):

$$z_2=\begin{bmatrix}0\\ 0\\ \boldsymbol{m}_3\\ \boldsymbol{m}_4\end{bmatrix}=h(\boldsymbol{x}_k,\boldsymbol{m}_1,\boldsymbol{m}_2,\boldsymbol{m}_3,\boldsymbol{m}_4)=\begin{bmatrix}\tilde{\boldsymbol{m}}_3^T\boldsymbol{R}_{1,3}^T\left[\boldsymbol{t}_{1,3}\times\right]\tilde{\boldsymbol{m}}_1\\ \tilde{\boldsymbol{m}}_4^T\boldsymbol{R}_{2,4}^T\left[\boldsymbol{t}_{2,4}\times\right]\tilde{\boldsymbol{m}}_2\\ \boldsymbol{K}_1\left(\sum_i\tilde{\boldsymbol{m}}_{1i}{}^1\boldsymbol{T}_i^T\right)l_2\\ \boldsymbol{K}_2\left(\sum_i\tilde{\boldsymbol{m}}_{1i}{}^2\boldsymbol{T}_i^T\right)l_2\end{bmatrix} \tag{13}$$

The measurement models using the epipolar constraint, the trifocal tensor constraint and the quadrifocal tensor constraint can be described as:

$$z=\begin{bmatrix}z_1\\ z_2\end{bmatrix} \tag{14}$$

In Equations (13) and (14), $\{\boldsymbol{m}_1,\boldsymbol{m}_2,\boldsymbol{m}_3,\boldsymbol{m}_4\}$ are the feature points corresponding to the four images; $\tilde{\boldsymbol{m}}_{1i}$ is the i-th element of vector $\tilde{\boldsymbol{m}}_1$, $\tilde{\boldsymbol{m}}_{2j}$ is the j-th element of vector $\tilde{\boldsymbol{m}}_2$, $\tilde{\boldsymbol{m}}_{3k}$ is the k-th element of vector $\tilde{\boldsymbol{m}}_3$, $\tilde{\boldsymbol{m}}_{4l}$ is the l-th element of vector $\tilde{\boldsymbol{m}}_4$; $\tilde{\boldsymbol{m}}_1$ equals $\boldsymbol{K}_1^{-1}\boldsymbol{m}_1$, $\tilde{\boldsymbol{m}}_2$ equals $\boldsymbol{K}_2^{-1}\boldsymbol{m}_2$, $\tilde{\boldsymbol{m}}_3$ equals $\boldsymbol{K}_1^{-1}\boldsymbol{m}_3$ and $\tilde{\boldsymbol{m}}_4$ equals $\boldsymbol{K}_2^{-1}\boldsymbol{m}_4$. $\boldsymbol{p}_1^p,\boldsymbol{p}_2^q,\boldsymbol{p}_3^r,\boldsymbol{p}_4^s$ denote the p-th, q-th, r-th and s-th row of the projection matrices $\boldsymbol{P}_1,\boldsymbol{P}_2,\boldsymbol{P}_3,\boldsymbol{P}_4$, respectively. $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ are

https://doi.org/10.1017/S0373463318000310 Published online by Cambridge University Press

the camera intrinsic parameter matrices of the left and right cameras, respectively. ${}^{1}\boldsymbol{T}$ and ${}^{2}\boldsymbol{T}$ are the trifocal tensors of the views (Im1, Im2, Im3) and (Im1, Im2, Im4), respectively. $\left[\boldsymbol{t}_{j,j+2}\times\right]$ and $\boldsymbol{R}_{j,j+2}$ are the position and rotation matrix of the frame C_{j+2} with respect to the frame C_j, respectively. The lines $\boldsymbol{l}_2, l_4$ are recommended choices as the lines perpendicular to the epipolar lines (Hartley et al., 2003).

The tensor ε_{ijk} is defined as (Hartley et al., 1998):

$$\varepsilon_{ijk} = \begin{cases} 0, & \text{unless } i,j,k \text{ are distinct} \\ +1, & \text{if } i,j,k \text{ are an even permutation of } 1,2,3 \\ -1, & \text{if } i,j,k \text{ are an odd permutation of } 1,2,3 \end{cases} \tag{15}$$

The rotational and translational relationship between the IMU and different views of the Camera (which is used in Equations (12) and (13)) are derived below.

We denote that $\boldsymbol{T}_{IW}$ and ${}^{1}\boldsymbol{T}_{IW}$ are the transformation matrices from frame $\{W\}$ to frame $\{I\}$ at the current time and last time, respectively; $\boldsymbol{T}_{CI}$ are the transformation matrices from frame $\{I\}$ to frame $\{C_L\}$; $\boldsymbol{T}_{C1W}$ are the transformation matrices from frame $\{W\}$ to first view camera frame $\{C_1\}$; $\boldsymbol{T}_{C3W}$ are the transformation matrices from frame $\{W\}$ to third view camera frame $\{C_3\}$; $\boldsymbol{R}_{C1W}$ and $\boldsymbol{R}_{C3W}$ are the rotation matrices from frame $\{W\}$ to $\{C_1\}$ and $\{C_3\}$, respectively; $\boldsymbol{t}_{WC1}$ is the translational vector from $\{C_1\}$ to $\{W\}$ with respect to $\{C_1\}$ and $\boldsymbol{t}_{WC3}$ is the translational vector from $\{C_3\}$ to $\{W\}$ with respect to $\{C_3\}$.

Then $\boldsymbol{T}_{IW}$ and ${}^{1}\boldsymbol{T}_{IW}$ can be expressed as:

$$\begin{aligned} \boldsymbol{T}_{IW} &= \begin{bmatrix} \left(\boldsymbol{R}(\bar{\boldsymbol{q}}_{WI})\right)^T & -\left(\boldsymbol{R}(\bar{\boldsymbol{q}}_{WI})\right)^T \boldsymbol{p}_{WI}^W \\ 0 \quad 0 \quad 0 & 1 \end{bmatrix} \\ {}^{1}\boldsymbol{T}_{IW} &= \begin{bmatrix} \left(\boldsymbol{R}\left({}^{1}\left(\bar{\boldsymbol{q}}_{WI}\right)\right)\right)^T & -\left(\boldsymbol{R}\left({}^{1}\left(\bar{\boldsymbol{q}}_{WI}\right)\right)\right)^T {}^{1}\left(\boldsymbol{p}_{WI}^W\right) \\ 0 \quad 0 \quad 0 & 1 \end{bmatrix} \end{aligned} \tag{16}$$

with:

$$\begin{aligned} \boldsymbol{T}_{C1W} &= \boldsymbol{T}_{CI}^{1}\boldsymbol{T}_{IW} & \boldsymbol{T}_{C3W} &= \boldsymbol{T}_{CI}\boldsymbol{T}_{IW} \\ \boldsymbol{R}_{C1W} &= \boldsymbol{T}_{C1W}(1:3,1:3) & \boldsymbol{R}_{C3W} &= \boldsymbol{T}_{C3W}(1:3,1:3) \\ \boldsymbol{t}_{WC1} &= \boldsymbol{T}_{C1W}(1:3,4) & \boldsymbol{t}_{WC3} &= \boldsymbol{T}_{C3W}(1:3,4) \end{aligned} \tag{17}$$

Therefore, the rotation matrices and translational vectors among different views can be derived as:

$$\begin{aligned} \boldsymbol{R}_{31} &= \boldsymbol{R}_{C3W}(\boldsymbol{R}_{C1W})^T & \boldsymbol{t}_{43} &= \boldsymbol{t}_{21} \\ \boldsymbol{R}_{43} &= \boldsymbol{R}_{21} & \boldsymbol{t}_{31} &= \boldsymbol{t}_{WC1} - (\boldsymbol{R}_{31})^T \boldsymbol{t}_{WC3} \\ \boldsymbol{R}_{41} &= \boldsymbol{R}_{43}\boldsymbol{R}_{31} & \boldsymbol{t}_{41} &= \boldsymbol{t}_{31} + (\boldsymbol{R}_{31})^T \boldsymbol{t}_{43} \\ \boldsymbol{R}_{42} &= \boldsymbol{R}_{41}(\boldsymbol{R}_{21})^T & \boldsymbol{t}_{42} &= \boldsymbol{R}_{21}(\boldsymbol{t}_{41} - \boldsymbol{t}_{21}) \end{aligned} \tag{18}$$

It can be seen from Equations (12) and (13) that the measurement models are dependent because the feature point inputs for them are the same. The measurement noise at the current time is also correlated with the measurement noise at the previous time. So, the performance of the proposed method can be further improved when such correlation is utilised. An example of this is by using nonlinear Kalman filters (Huang et al., 2016).

https://doi.org/10.1017/S0373463318000310 Published online by Cambridge University Press

However, it is very difficult to obtain an accurate model for such correlation of measurement noise in engineering applications. In this paper, to facilitate the design of a nonlinear filter, the measurement noise is modelled as white noise. Fortunately, the proposed method exhibits satisfactory performance based on such a model, as will be shown in experiment.

4. OUTLIER REJECTION. For feature detection, the fast corner detection algorithm (FAST) is used (Rosten et al., 2006). For feature tracking, the Kanade Lucas Tomasi (KLT) tracker is used (Shi, 1994), which allows tracking features over long image sequences. For the outlier rejection, the epipolar constraint, the trifocal tensor constraint and the quadrifocal tensor constraint are used.

The epipolar constraint is designed to remove the outlier between left and right images, for example, the constraint between $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ which is shown in Equation (19):

$$\{\boldsymbol{m}_1, \boldsymbol{m}_2\}_{inliers} = \left\{ \{\boldsymbol{m}_1, \boldsymbol{m}_2\} \,\middle|\, \left\| \tilde{\boldsymbol{m}}_2^T R_{21} \left[t_{21} \times \right] \tilde{\boldsymbol{m}}_1 \right\| < th_{epi} \right\} \tag{19}$$

where th_{epi} is the threshold for the constraint.

The RANdom SAmple Consensus (RANSAC) algorithm is then used by using the trifocal tensor constraint and the quadrifocal tensor constraint which are displayed in Equations (20) and (21).

$$\{\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3\}_{inliers} = \left\{ \{\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3\} \,\middle|\, \left\| \left(\boldsymbol{l}_2^T \left(\sum_i \tilde{\boldsymbol{m}}_{1i}{}^1 \boldsymbol{T}_i^T \right) \left[\tilde{\boldsymbol{m}}_3 \times \right] \right)^T \right\| < th_{tri} \right\} \tag{20}$$

$$\{\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4\}_{inliers} = \left\{ \{\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4\} \,\middle|\, \left\| \tilde{\boldsymbol{m}}_{1i} \tilde{\boldsymbol{m}}_{2j} \tilde{\boldsymbol{m}}_{3k} \boldsymbol{l}_{4s} \varepsilon_{ipw} \varepsilon_{jqx} \varepsilon_{kry} Q^{pqrs} \right\| < th_{quadri} \right\} \tag{21}$$

where th_{tri} and th_{quadri} are the thresholds for the trifocal tensor constraint and the quadrifocal tensor constraint, respectively.

5. MINIMUM SIGMA SET SQUARE-ROOT UNSCENTED KALMAN FILTER. Minimum sigma set unscented transformation was proposed by Menegaz (2011) by using only $n + 1$ sigma points (both simplex UT and spherical simplex UT use $n + 2$ sigma points) to match the mean and the covariance matrices of the system state.

The Kalman filter flow of the M-SRUKF used in the vision-IMU tightly-coupled system is shown in Figure 3.

6. EXPERIMENTS. In order to examine the performance of the proposed algorithms, two open source datasets were used. The data in the first dataset was collected from a land vehicle, while the data in the second dataset was collected from a MAV.

6.1. *Land vehicle test.* The first dataset is from the Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) Vision Benchmark Suite (Geiger et al., 2012). In all the experiments, the sensors were rigidly mounted on top of a land vehicle. The intrinsic parameters of the cameras and the transformation between the cameras and GPS/IMU were well calibrated. Moreover, the cameras and GPS/IMU were manually synchronised, with sampling rates of 10 Hz and 100 Hz, respectively. The source data used here are the un-synchronised IMU measurements sampled at 100 Hz and the rectified grayscale stereo

https://doi.org/10.1017/S0373463318000310 Published online by Cambridge University Press

1) Initial values of the state and the state covariance:

$$\boldsymbol{x}_{0|0}=\bar{\boldsymbol{x}}_0\,,\ \boldsymbol{S}_{\boldsymbol{x}_{0|0}}=chol\left\{E[(\boldsymbol{x}_{0|0}-\hat{\boldsymbol{x}}_0)(\boldsymbol{x}_{0|0}-\hat{\boldsymbol{x}}_0)^T]\right\}$$

2) Initial coefficients for M-SRUKF:

Choose $0\le W_0\le 1$, and set $\rho:=\sqrt{(1-W_0)/n}$, $C:=\sqrt{I_n-\rho^2[1]_{n\times n}}$, $W=diag(W_1,...,W_n)$. For $i=1,...,n: W_i=[C^{-1}W_0\rho^2[1]_{n\times n}(C^T)^{-1}]_{i,i}$, which is the weight that associates with the *i*-th sigma point.

3) Time update equations:

Propagate directly through state transition function

$\boldsymbol{x}_{k|k-1}=\boldsymbol{F}_k\boldsymbol{x}_{k-1|k-1}, \boldsymbol{S}_{\boldsymbol{x}_{k|k-1}}=\boldsymbol{F}_k\boldsymbol{S}_{\boldsymbol{x}_{k-1|k-1}}, \boldsymbol{S}_{\boldsymbol{x}_{k|k-1}}=qr(\boldsymbol{S}_{\boldsymbol{x}_{k|k-1}}\ \ \boldsymbol{Q}^{1/2})$, where $\boldsymbol{F}_k$ is the systematic matrix after discretisation. The state of the filter is the form of the error, so the mean and covariance can be predicted directly without using the spreading sigma points. This may greatly save computation time. $\boldsymbol{S}_{\boldsymbol{x}_{k|k-1}}=\boldsymbol{F}_k\boldsymbol{S}_{\boldsymbol{x}_{k-1|k-1}}$ is derived from $\boldsymbol{P}_{\boldsymbol{x}_{k|k-1}}=\boldsymbol{F}_k\boldsymbol{P}_{\boldsymbol{x}_{k-1|k-1}}\boldsymbol{F}_k'$, where $\boldsymbol{P}_{\boldsymbol{x}_{k-1|k-1}}$ is the covariance matrix of the state. $qr(\cdot)$ represents QR decomposition function.

4) Calculate Sigma points:

$$\boldsymbol{\chi}=[\chi_0\ \ ...\ \ \chi_n]:=[-\boldsymbol{S}_{\boldsymbol{x}_{k-1|k-1}}[\alpha]_{n\times 1}/\sqrt{W_0}\ \ \boldsymbol{S}_{\boldsymbol{x}_{k-1|k-1}}C(\sqrt{W})^{-1}]+[\boldsymbol{x}_{k-1|k-1}]_{1:n+1},$$

The column vectors of $\boldsymbol{\chi}$ are the sigma point vectors, k is the iteration time starts from 1.

5) Measurement update equations: $\boldsymbol{\xi}'=h(\boldsymbol{\chi})$.

Calculate Mean of the predicted measurements: $\boldsymbol{y}_k'=\sum_{i=0}^{n}W_i\boldsymbol{\xi}_i'$

Calculate the covariance value of the measurements:

$\boldsymbol{S}_{y,k}'=qr(\sqrt{W_i}\ (\boldsymbol{\xi}_{1:2n}'-\boldsymbol{y}_k')\ \ \boldsymbol{R}^{1/2}), \boldsymbol{S}_{y,k}'=cholupdate(\boldsymbol{S}_{y,k}', \boldsymbol{\xi}_0'-\boldsymbol{y}_k', W_0)$

Update state and state covariance: $\boldsymbol{P}_{xz,k}=\sum_{i=0}^{n}W_i\ (\boldsymbol{\chi}_i-\boldsymbol{x}_{k|k-1})(\boldsymbol{\xi}_i'-\boldsymbol{y}_k')^T$,

$\boldsymbol{K}_k=\boldsymbol{P}_{xz,k}/\boldsymbol{S}_{y,k}'/(\boldsymbol{S}_{y,k}')^T, \boldsymbol{x}_{k|k}=\boldsymbol{x}_{k|k-1}+\boldsymbol{K}_k(\boldsymbol{y}_k-\boldsymbol{y}_k'), \ U=\boldsymbol{K}_k\boldsymbol{S}_{y,k}'$.

Update the Cholesky factors via rank-one Cholesky updates:

$\boldsymbol{S}_{\boldsymbol{x}_{k|k}}=cholupdate(\boldsymbol{S}_{\boldsymbol{x}_{k-1|k-1}}, U, -1)$, which will be used in the next step 3.

Figure 3. M-SRUKF flow used in the vision-IMU tightly-coupled system.

https://doi.org/10.1017/S0373463318000310 Published online by Cambridge University Press

Table 2. Horizontal position errors of the land vehicle test.

Algorithms	Position RMSE (m)	Position MAXE (m)	Position ENDE (m)
ThreeViews-SRUKF-Stereo/IMU	11·37	22·01	19·39
FourViews-M-SRUKF-Stereo/IMU1	8·44	22·60	18·57
FourViews-M-SRUKF-Stereo/IMU2	**4·90**	**14·91**	**11·73**

sequences sampled at 10 Hz. The ground truth was taken from a GPS/IMU integrated solution with Real-Time Kinematic (RTK) correction signals (OXTS RT3003). The gyroscope and accelerometer bias specifications are 36°/h (1σ) and 1 mg (1σ), respectively. The resolution of stereo images is 1226 × 370 pixels, with a 90° field view. In the testing results presented in the following parts, the information below is considered:

- The proposed algorithms were developed and tested using the MATLAB R2016b on a 64-bit operating system, with an Intel Core i7-6700k CPU @ 4.00GHz, and the installed memory was 16.0 GB.
- The affection of initial weight W_0 is considered here, which is set to 0.5 for the M-SRUKF.
- The symbols used in this paper are defined as follows:
  - ThreeViews-SRUKF-Stereo/IMU: SRUKF-based binocular stereo vision aided inertial navigation solution by using trifocal constraints (Xian et al., 2016).
  - FourViews-M-SRUKF-Stereo/IMU1: M-SRUKF-based binocular stereo vision aided inertial navigation solution by using the four views constraints in Equation (14), with only the quadrifocal tensor constraint $0_{123} = \tilde{\boldsymbol{m}}_{1i}\tilde{\boldsymbol{m}}_{2j}\tilde{\boldsymbol{m}}_{3k}\boldsymbol{l}_{4s}\varepsilon_{ip1}\varepsilon_{jq2}\varepsilon_{kr3}\mathcal{Q}^{pqrs}$ in $\boldsymbol{z}_1$ used.
  - FourViews-M-SRUKF-Stereo/IMU2: M-SRUKF-based binocular stereo vision aided inertial navigation solution by using the four views constraints in Equation (14), with all the quadrifocal tensor constraints in $\boldsymbol{z}_1$ used.

The test chosen has a trajectory of about 4.1 km (which is the longest trajectory in the dataset) and lasts 518 s with average velocity of 8 m/s. The horizontal position errors are displayed in Table 2. Note that the large error from 75 s to 120 s is caused by ground truth error due to GPS signal loss. For this reason, this period was not taken into consideration when calculating the overall Root Mean Square Error (RMSE), the Maximum Error (MAXE) and the End point Error (ENDE). Figure 4(a) and 4(b) show the estimated trajectories of different algorithms on a Google Map background and the corresponding horizontal position errors with time.

The results clearly show that the FourViews-M-SRUKF-Stereo/IMU2 has the highest positioning accuracy, with overall position error RMSE 4.90 m, MAXE 14.91 m and ENDE 11.73 m. However, the ThreeViews-SRUKF-Stereo/IMU has overall position error RMSE 11.37 m, MAXE 22.01 m and ENDE 19.39 m. The FourViews-M-SRUKF-Stereo/IMU2 has improved the accuracy by adding the quadrifocal tensor constraints. The FourViews-M-SRUKF-Stereo/IMU1 also shows higher accuracy than the ThreeViews-SRUKF-Stereo/IMU in terms of position error RMSE, but both systems have a similar MAXE accuracy. The FourViews-M-SRUKF-Stereo/IMU2 has higher accuracy than the FourViews-M-SRUKF-Stereo/IMU1. This is because the former uses more quadrifocal

https://doi.org/10.1017/S0373463318000310 Published online by Cambridge University Press

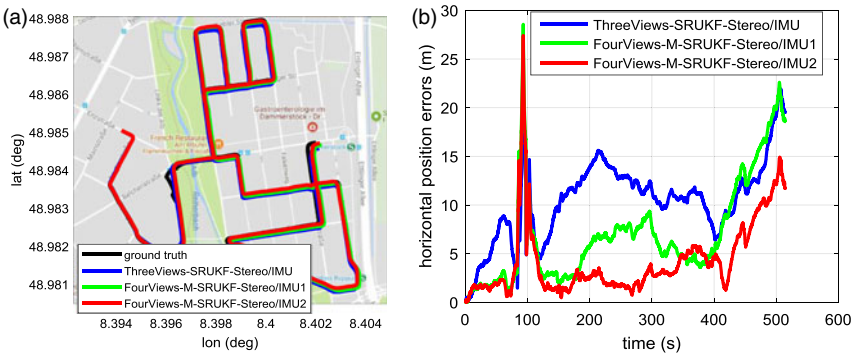


Figure 4. Estimated trajectories of different algorithms of the land vehicle test. (a) Trajectories on google map. (b) Horizontal positions errors of different algorithms.

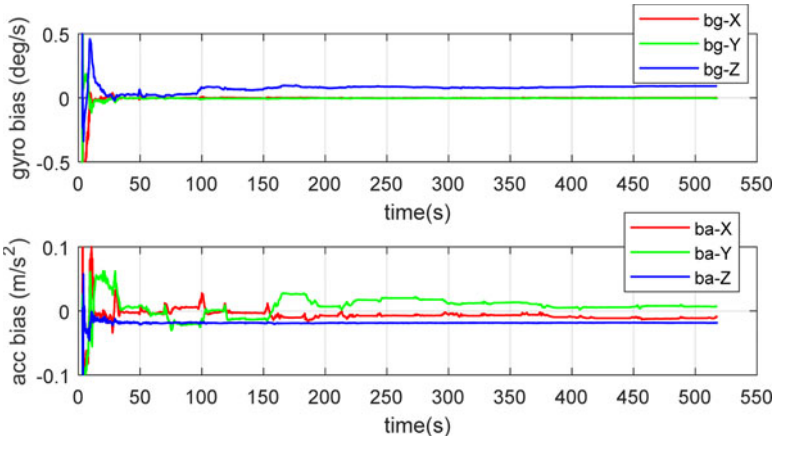


Figure 5. Estimated gyro and accelerometer biases of the land vehicle test.

tensor constraints than the latter, which is more reliable for the Kalman filter measurements and the outlier rejection process.

Figure 5 indicates that the new binocular vision-IMU mechanisation can estimate the gyro and the accelerometer biases successfully, which is crucial for long endurance autonomous navigation.

A computation load comparison among the ThreeViews-SRUKF-Stereo/IMU, the FourViews-M-SRUKF-Stereo/IMU1 and the FourViews-M-SRUKF-Stereo/IMU2 is also conducted. The computation time (by using MATLAB) of the measurement updating process and the outlier rejection process are compared.

The major differences of the two FourViews-M-SRUKF-Stereo/IMUs and the ThreeViews-SRUKF-Stereo/IMU are that the FourViews-M-SRUKF-Stereo/IMUs utilise the M-SRUKF flow which is using only $n + 1$ sigma points, while the ThreeViews-SRUKF-Stereo/IMU uses the standard SRUKF flow which uses $2n + 1$ sigma points; the former has nearly half the computation load of the latter. However, the FourViews-M-SRUKF-Stereo/IMUs has more computation load on the Kalman filter measurements and the outlier rejection process, which can be seen from Equations (12) and (21).

https://doi.org/10.1017/S0373463318000310 Published online by Cambridge University Press

Table 3. Computation cost comparison.

Algorithms	Measurement Updating	Outlier Rejection
ThreeViews-SRUKF-Stereo/IMU	96·484 s	66·457 s
FourViews-M-SRUKF-Stereo/IMU1	105·958 s	76·107 s
FourViews-M-SRUKF-Stereo/IMU2	168·682 s	115·150 s

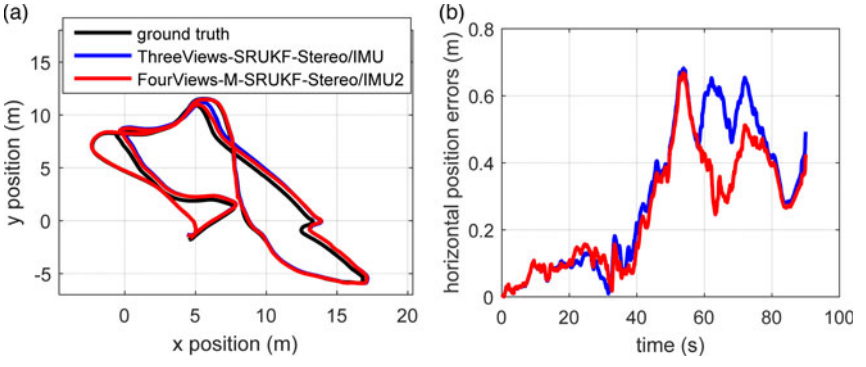


Figure 6. Estimated trajectories of different algorithms of the MAV test. (a) Top-down view of the trajectories. (b) Horizontal positions errors of different algorithms.

The measurement updating time and outlier rejection time in Table 3 show that both the FourViews-M-SRUKF-Stereo/IMU1 and the FourViews-M-SRUKF-Stereo/IMU2 have more computation cost than the ThreeViews-SRUKF-Stereo/IMU. However, in real time application by using C++, this is not a problem. Though the computation load of the FourViews-M-SRUKF-Stereo/IMU2 is heaviest, the accuracy improvement is promising.

6.2. *MAV test.* The second dataset is from the European Robotics Challenge (EUROC) which covers the machine hall at Eidgenössische Technische Hochschule (ETH) (Burri et al., 2016): a typical environment to demonstrate industrial inspection using MAV. The stereo images (752×480) are 20 Hz, and IMU (ADIS16448, micro electro mechanical systems-MEMS) data is 200 Hz, they were all precisely synchronised. The ground truth was obtained by using a laser-tracker, which precisely measured the Three-Dimensional (3D) position by tracking a prism mounted on the MAV. A batch optimisation is employed offline, thus the six Degree of Freedom (DoF) pose in the ground truth was also provided.

A typical case from the dataset is chosen for algorithm evaluation, which was "MH_05_diffcult". The initial inputs of our algorithms were set from the ground truth data. The position result comparison among different algorithms is shown in Figure 6.

Figure 7 indicates that the new binocular vision-IMU mechanisation can estimate the gyro and the accelerometer biases successfully in the MAV test, because all the biases converge quickly to nearly constant values.

All the land vehicle and MAV test results show that the new mechanisation FourViews-M-SRUKF-Stereo/IMU2 has the best accuracy, which means that the mechanisation ThreeViews-SRUKF-Stereo/IMU by merely using the epipolar and the trifocal tensor constraints in Equation (13) is not enough to cover all the useful measurement information.

https://doi.org/10.1017/S0373463318000310 Published online by Cambridge University Press

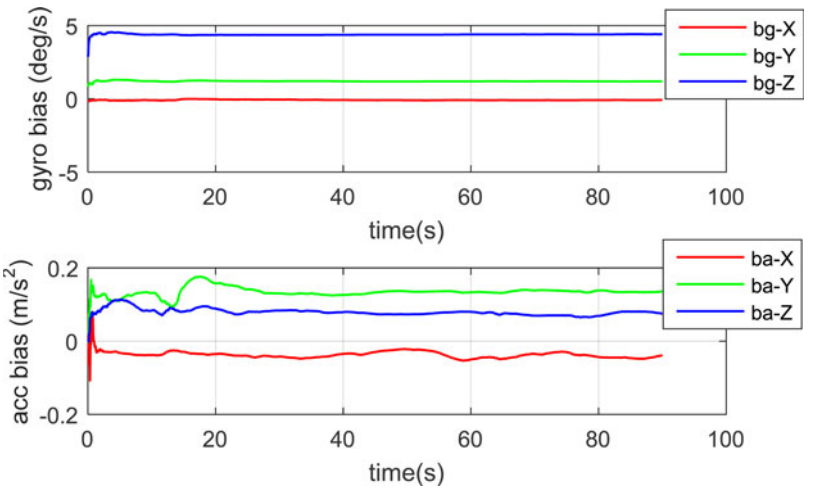


Figure 7. Estimated gyro and accelerometer biases of the MAV test.

However, the quadrifocal tensor constraint has the potential to provide more accurate measurement information which is not implied in both the epipolar and the trifocal tensor constraints. The quadrifocal tensor constraint has not only been used in the Kalman filter measurement model, but also used in the outlier rejection process. This implies that the extracted features are more accurate due to the addition of the quadrifocal tensor geometry constraint among four views. So, the proposed quadrifocal tensor mechanisation is an effective method to be used in the binocular vision-IMU tightly-coupled system.

7. CONCLUSIONS. In this paper, we propose a binocular vision-IMU tightly-coupled structure based on M-SRUKF for real time navigation. The reason for choosing M-SRUKF is that it has only half the sigma points of a standard SRUKF while their accuracy are nearly the same for a binocular vision-IMU tightly-coupled system. The measurement model and the outlier rejection model not only utilised the epipolar constraint and the trifocal tensor geometry constraints between the consecutive two image pairs, but also used the quadrifocal tensor geometry among four views between the consecutive times.

The experiments were carried out by using an open source KITTI dataset and a MAV dataset. Results showed that the new mechanisation (FourViews-M-SRUKF-Stereo/IMU2) worked well in both cases. In the longest land vehicle trajectory, the overall position error RMSE and MAXE were 4.90 m and 14.91 m, respectively, which is better than 11.37 m and 22.01 m by using the ThreeViews-SRUKF-Stereo/IMU. The typical case in the MAV test also shows the high position accuracy of the new mechanisation. The proposed method can greatly suppress the navigation error, so it is suitable for real time navigation, especially for GPS signal-denied environments.

ACKNOWLEDEMENT

The authors want to express their sincere thanks for the support of China Scholarship Council (CSC, File No. 201603170338).

https://doi.org/10.1017/S0373463318000310 Published online by Cambridge University Press

REFERENCES

Asadi, E. and Bottasso, C. L. (2012). Tightly-coupled vision-aided inertial navigation via trifocal constraints. *IEEE International Conference on Robotics and Biomimetics*, 85–90.

Asadi, E. and Bottasso, C. L. (2014). Tightly-coupled stereo vision-aided inertial navigation using feature-based motion sensors. *Advanced Robotics*, **28**(11), 717–729.

Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M. W. and Siegwart, R. (2016). The EUROC micro aerial vehicle datasets. *International Journal of Robotics Research*, **35**(10), 1157–1163.

Corke, P., Lobo, J. and Dias, J. (2007). An introduction to inertial and visual sensing. *International Journal of Robotics Research*, **26**(6), 519–535.

Fang, Q. and Huang, S. X. (2013). UKF for Integrated Vision and Inertial Sensors Based on Three-View Geometry. *IEEE Sensors Journal*, **13**(7), 2711–2719.

Faugeras, O. and Mourrain, B. (1995, June). On the geometry and algebra of the point and line correspondences between n images. *IEEE International Conference on Computer Vision* (pp. 951–956).

Feng, G., Wu, W. and Wang, J. (2012). Observability analysis of a matrix Kalman filter-based navigation system using visual/inertial/magnetic sensors. *Sensors*, **12**(7), 8877–8894.

Geiger, A., Lenz, P. and Urtasun, R. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. *IEEE Conference on Computer Vision and Pattern Recognition*, 3354–3361.

Hartley, R. (1998). Computation of the quadrifocal tensor. *European Conference on Computer Vision*, 20–35.

Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge University Press.

Hesch, J. A., Kottas, D. G., Bowman, S. L. and Roumeliotis, S. I. (2014). Camera-IMU-based localization: Observability analysis and consistency improvement. *The International Journal of Robotics Research*, **33**(1), 182–201.

Hu, J. S. and Chen, M. Y. (2014). A sliding-window visual-IMU odometer based on tri-focal tensor geometry. *IEEE International Conference on Robotics and Automation* (pp. 3963–3968).

Huang, Y., Zhang, Y., Li, N. and Chambers, J. (2016). A robust Gaussian approximate fixed-interval smoother for nonlinear systems with heavy-tailed process and measurement noises. *IEEE Signal Processing Letters*, **23**(4), 468–472.

Huang, Y., Zhang, Y., Li, N. and Chambers, J. (2016). Robust Student'st based nonlinear filter and smoother. *IEEE Transactions on Aerospace and Electronic Systems*, **52**(5), 2586–2596.

Indelman, V., Gurfil, P., Rivlin, E., Rotstein, H. (2012). Real-Time Vision-Aided Localization and Navigation Based on Three-View Geometry. *IEEE Transactions on Aerospace and Electronic Systems*, **48**(3), 2239–2259.

Julier, S. J. (2003). The spherical simplex unscented transformation. *Proceedings of the 2003 American Control Conference* (pp. 2430–2434).

Julier, S. J. and Uhlmann, J. K. (2002). Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations. *Proceedings of the 2002 American Control Conference* (pp. 887–892).

Kelly, J. and Sukhatme, G. S. (2011). Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration. *The International Journal of Robotics Research*, **30**(1), 56–79.

Kitt, B., Geiger, A. and Lategahn, H. (2010). Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme. *IEEE Intelligent Vehicles Symposium* (pp. 486–492).

Kong, X., Wu, W., Zhang, L. and Wang, Y. (2015). Tightly-coupled stereo visual-inertial navigation using point and line features. *Sensors*, **15**(6), 12816–12833.

Kottas, D. G. and Roumeliotis, S. I. (2013). Efficient and consistent vision-aided inertial navigation using line observations. *IEEE International Conference on Robotics and Automation* (pp. 1540–1547).

Li, W. C., Wei, P., and Xiao, X. C. (2007). A novel simplex unscented transform and filter. *IEEE International Symposium on Communications and Information Technologies*, 61–65.

Liu, F., Sarvrood, Y. B. and Gao, Y. (2018). Implementation and Analysis of Tightly Integrated INS/Stereo VO for Land Vehicle Navigation. *The Journal of Navigation*, **71**(1), 83–99.

Menegaz, H. M. T., Ishihara, J. Y., Borges, G. A., and Vargas, A. N. (2015). A Systematization of the Unscented Kalman Filter Theory. *IEEE Transactions on Automatic Control*, **60**(10), 2583–2598.

Menegaz, H. M., Ishihara, J. Y. and Borges, G. A. (2011). A new smallest sigma set for the Unscented Transform and its applications on SLAM. *IEEE Conference on Decision and Control and European Control Conference*.

Merwe, R. V. d. and Wan, E. A. (2001). The square-root unscented Kalman filter for state and parameter-estimation. *International Conference on Acoustics, Speech, and Signal Processing* (pp. 3461–3464).

Mourikis, A. I. and Roumeliotis, S. I. (2007). A multi-state constraint Kalman filter for vision-aided inertial navigation. *IEEE international conference on Robotics and automation* (pp. 3565–3572).

https://doi.org/10.1017/S0373463318000310 Published online by Cambridge University Press

Rosten, E., and Drummond, T. (2006). Machine learning for high-speed corner detection. *European Conference on Computer Vision*, 430–443.

Shi, J. (1994). Good features to track. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 593–600.

Sirtkaya, S., Seymen, B., and Alatan, A. A. (2013). Loosely coupled Kalman filtering for fusion of visual odometry and inertial navigation. *International Conference on Information Fusion* (pp. 219–226).

Tardif, J. P., George, M., Laverne, M., Kelly, A., and Stentz, A. (2010). A new approach to vision-aided inertial navigation. *IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 4161–4168).

Wang M., Wu W. and El-sheimy, N. (2017). Comparisons of SR-UKF family for a Visual-IMU Tightly-coupled System Based on Tri-focal Tensor Geometry. *Proceedings of the ION GNSS+*, 3088–3101.

Weiss, S., Achtelik, M. W., Lynen, S., Chli, M. and Siegwart, R. (2012). Real-time onboard visual-inertial state estimation and self-calibration of MAVs in unknown environments. *IEEE International Conference on Robotics and Automation*, 957–964.

Xian, Z., Hu, X. and Lian, J. (2015). Fusing Stereo Camera and Low-Cost Inertial Measurement Unit for Autonomous Navigation in a Tightly-Coupled Approach. *The Journal of Navigation*, **68**(3), 434–452.

Xian, Z., Lian, J., Shan, M., Zhang, L., He, X. and Hu, X. (2016). A square root unscented Kalman filter for multiple view geometry based stereo cameras/inertial navigation. *International Journal of Advanced Robotic Systems,* **13**(5), 1–11.

Zhang, L. (2013). *Line primitives and their applications in geometric computer vision*. Universitätsbibliothek, Kiel.

https://doi.org/10.1017/S0373463318000310 Published online by Cambridge University Press