



RESEARCH ARTICLE

Stereo visual odometry with velocity constraint for ground vehicle applications

Fei Liu,^{1*} Yashar Balazadegan Sarvrood,² Yue Liu,³ and Yang Gao²

¹Department of Civil Engineering, University of Calgary, Calgary, Alberta, Canada.

²Department of Geomatics Engineering, University of Calgary, Calgary, Alberta, Canada.

³Department of Automation, Harbin Engineering University, Harbin, Heilongjiang, China.

*Corresponding author. E-mail: liu19@ucalgary.ca

Received: 21 June 2020; Accepted: 04 March 2021; First published online: 30 March 2021

Keywords: NHC, ground vehicle, stereo visual odometry

Abstract

This paper proposes a novel method of error mitigation for stereo visual odometry (VO) applied in land vehicles. A non-holonomic constraint (NHC), which imposes physical constraint to the rightward velocity of a land vehicle, is implemented as an observation in an extended Kalman filter (EKF) to reduce the drift of stereo VO. The EKF state vector includes position errors in an Earth-centred, Earth-fixed (ECEF) frame, velocity errors in the camera frame, angular rate errors and attitude errors. All the related equations are described and presented in detail. In this approach, no additional sensors are used but NHC, namely velocity constraint in the right direction, is applied as an external measurement to improve the accuracy. Tests are conducted with the Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) datasets. Results show that the relative horizontal positioning error improved from 0.63% to 0.22% on average with the application of the velocity constraints. The maximum and root mean square of the horizontal error with velocity constraints are both reduced to less than half of the error with stand-alone stereo VO.

1. Introduction

The global navigation satellite system (GNSS) is widely applied in autonomous driving and vehicle navigation. However, the performance of GNSS degrades in challenging environments such as dense urban areas, tunnels, or underground parking lots due to GNSS signal blockage and reflection (Groves and Jiang, 2013). Dead reckoning (DR) systems such as inertial navigation system (INS) or stereo visual odometry (VO) are widely used as systems complementary to GNSS in GNSS-denied environments (Liu, 2018; F. Liu et al., 2018a, 2018b). However, DR systems accumulate errors fast without external information. To reduce the accumulation of errors in DR systems, different sensors are integrated to reduce the rapidity of drift and increase system reliability.

The inertial measurement unit (IMU) is widely used in multi-sensor integration applications as the core sensor since it works in most circumstances (Hartley and Zisserman, 2001; Balazadegan Sarvrood, 2016; F. Liu et al., 2018a, 2018b). However, the main drawback of IMU is quick error accumulation without external information update. The results obtained from a low-cost micro-electrotechnical system (MEMS) IMU manifest drift in a short time (Jekeli, 2012; Noureldin et al., 2012). In theory, any self-dependent system can be deployed as the core component in an integrated navigation system. Stereo VO is a self-dependent DR system that is increasingly being utilised as the core sensor in multi-sensor fusion. For VO, the monocular camera is an orientation device only since the translation is estimated up to a scale ambiguity (Zhang et al., 2014). Adding another camera means that depth information is

achievable by the triangulation of stereo cameras (Scaramuzza and Fraundorfer, 2011). With VO, drift is inevitable because of error accumulation. To reduce the VO drift, feature matching or tracking is of significance in pose estimation. Random sample consensus (RANSAC) is applied in some studies to filter out incorrect features in estimation (Scaramuzza, 2011; Geiger et al., 2013; Liu et al., 2017). Key-frame-based bundle adjustment optimisation (Strasdat et al., 2012; Campos et al., 2020) and loop detection and relocalisation (Gálvez-López and Tardos, 2012; Campos et al., 2020) are also applied to improve the VO performance. By application of effective feature detection and matching, key-frame based optimisation, loop detection, etc., the accuracy and robustness of VO has been greatly improved (Cvišić and Petrović, 2015; Forster et al., 2016; Bloesch et al., 2017; Wang et al., 2017). In addition, an integrated system with other types of sensors (e.g., IMU) is another common way to provide external information to limit the drift (Bloesch et al., 2015; Leutenegger et al., 2015; Qin et al., 2018; Von Stumberg et al., 2018; Qin et al., 2019). In the past, little work has been devoted to stereo VO as the core sensor in integrated navigation. This paper presents the mechanisation equations of stereo VO and the corresponding linearised error equations which can be applied as the core component in integration with other types of measurements. For land vehicle applications, velocity constraint is an effective way of reducing the rapidity of drift based on the fact that the land vehicle only has forward velocity in most cases. To mitigate the error accumulation in land vehicle applications, velocity constraints in the right and down directions can be applied as additional measurements to reduce the stereo VO drift (Shin, 2001; Godha and Cannon, 2007; Noureldin et al., 2012; Y. Liu et al., 2018). Since the estimated state vector elements are correlated, the velocity constraints can correct the velocity estimation together with other unknowns. In this way, without adding any external sensors, the performance can be further improved for land vehicle applications. If external measurements are available, the proposed method can also be applied to integrate stereo VO with other systems.

The rest of this paper is organised as follows. In Section 2, the fundamentals of stereo VO are briefly reviewed. The method with stereo VO as the core component together with the NHC application in the extended Kalman filter (EKF) is illustrated in Section 3. Section 4 presents the test results of the proposed method using the KITTI datasets and the discussion and conclusions are in the final section.

2. Fundamentals of stereo VO

The principle of VO is to estimate the six degrees of freedom camera pose from consecutive images by solving perspective-n-point (PnP) (Fischler and Bolles, 1981). To estimate the unknown translation and rotation between the camera frames in two consecutive epochs, continuously tracked or matched image features together with their three-dimensional (3D) coordinates are applied in the resectioning (Nistér et al., 2004; Scaramuzza and Fraundorfer, 2011). Tracking or matching methods such as Kanade–Lucas–Tomasi (KLT) tracking, SIFT, FAST, SURF, etc. are used to find common features in consecutive images (Tomasi and Kanade, 1992; Shi and Tomasi, 1993; Bay et al., 2006; Rosten and Drummond, 2006). Several PnP methods such as P3P, EPnP and direct least squares are widely used for VO resectioning (Gao et al., 2003; Lepetit et al., 2009; Hesch and Roumeliotis, 2011).

In this paper, 3D feature coordinates are determined by the triangulation of stereo cameras. A KLT tracker is used to track the common features continuously in consecutive frames (Shi and Tomasi, 1993). The camera pose estimation is based on the least squares method proposed by Menze and Geiger (2015). The general procedures can be summarised as follows. With the feature tracker, the common features on stereo images in two consecutive epochs can be found. The translation and rotation unknowns between two consecutive epochs can be calculated by minimising the difference between the reprojected and tracked image coordinates of the features on the current stereo images in the least squares estimation.

The coordinate systems used in this paper are defined as follows. The origin of the camera frame is defined at the perspective centre of the left camera with the X -axis pointing to the right, the Y -axis pointing downward and the Z -axis completing the right-handed coordinate system. The origin of the image coordinate system is the left-top corner with X pointing to the right and Y pointing to the down direction. The coordinate conversion from the previous camera frame to the current camera frame can

be given as (Menze and Geiger, 2015)

$$C^c = R_p^c C^p + T_{cp}^c$$

$$R_p^c = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix} \begin{bmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

where C^c and C^p are the feature coordinates in the current and previous camera frames; R_p^c is the rotation matrix from the previous camera frame (p) to the current camera frame (c); γ, β, α are the Euler angles rotating with the Z -axis in the previous camera frame Z^p , Y -axis after the first rotation Y' and the X -axis after the second rotation X'' respectively; $T_{cp}^c = [tx, ty, tz]^T$ is the translation expressed in the current camera frame.

To reproject the feature coordinate $C^c (x_c, y_c, z_c)$ in the current camera frame on the image (u, v) , the following equation is applied (Scaramuzza and Fraundorfer, 2011)

$$u = f_x \frac{x_c}{z_c} + u_0$$

$$v = f_y \frac{y_c}{z_c} + v_0 \quad (2)$$

where (u, v) is the feature image coordinate; (x_c, y_c, z_c) is the 3D feature coordinate in the current camera frame; (u_0, v_0) is the principal point image coordinate; the focal length of image X - and Y -axes are represented as f_x and f_y .

To estimate the unknowns $[\alpha, \beta, \gamma, tx, ty, tz]^T$, initial values are set and the corrections to the initial values are estimated using the least squares equations. For six unknowns, at least three features are required to form six equations (Lepetit et al., 2009). The RANSAC algorithm is applied to remove the outliers (Nistér, 2005). Specifically, three features are randomly selected to estimate the rotation and translation 50 times. In each estimation, the number of inliers is recorded with a certain threshold. The maximum number of inliers is applied again to finally achieve the estimation, given as (Menze and Geiger, 2015)

$$dX = (H^T PH)^{-1} H^T PL \quad (3)$$

with estimated covariance matrix as

$$R = (H^T PH)^{-1} \quad (4)$$

In Equation (3), dX is the estimated corrections to the initial values of the unknowns $[\alpha, \beta, \gamma, tx, ty, tz]^T$; P is the weight matrix; L is the misclosure of the observed feature pixel coordinates minus the calculated ones achieved by Equations (1) and (2); H is the design matrix. In Equation (4), R is the covariance matrix of dX in Equation (3); P and H have the same meaning as in Equation (3). In normal cases, all the elements in R are expected to be non-zero, which means that the unknowns (translation and rotation) are correlated with each other.

3. Stereo VO as a core component

So far, the procedures for solving the translation and rotation of stereo VO between consecutive epochs have been introduced. By accumulating the rotation and translation, the camera pose at any epoch can be achieved. To integrate stereo VO with other types of measurements as a core component, the stereo VO error dynamic equations have to be derived to be applied in the Kalman filter. Therefore, another way to apply the stereo VO is to form the stereo VO mechanisation using the results from the previous section. In this section, the stereo VO mechanisation, error linearisation, Kalman filter integration and ground vehicle velocity constraint are illustrated.

3.1. Stereo VO mechanisation

To apply the stereo VO as a core component in integrated navigation, the mechanisation of stereo VO can be given as Equation (5), based on the results achieved by Equations (3) and (4). In this paper, the mechanisation of stereo VO is conducted in ECEF (Earth-centred, Earth-fixed), given as

$$\begin{bmatrix} \dot{r}^e \\ \dot{R}_c^e \end{bmatrix} = \begin{bmatrix} R_c^e v^c \\ R_c^e \Omega_{ec}^c \end{bmatrix} \tag{5}$$

where the left side of the equation is the time derivatives of the r^e and R_c^e ; r^e is the position in ECEF; R_c^e is the rotation matrix from camera frame to ECEF; Ω_{ec}^c is the skew-symmetric matrix of the angular rate of camera frame with respect to ECEF expressed in the camera frame ω_{ec}^c ; v^c is the velocity in the camera frame.

Note that the initial position, velocity and attitude are assumed to be known. Therefore, to apply Equation (5) for stereo VO navigation solutions, v^c and Ω_{ec}^c need to be calculated, which can be given as follows. From Equations (1) to (3), the translation in the camera frame can be achieved. The velocity is the translation divided by time interval, given as

$$v^c = T_{cp}^c / \Delta t \tag{6}$$

where $T_{cp}^c = [tx, ty, tz]^T$ is the translation obtained from Equation (3); Δt is the time interval between two consecutive epochs.

For the angular rate ω_{ec}^c , it can be calculated based on the Euler angles and their corresponding axis. As mentioned in Equation (1), the rotation can be decomposed as Euler angles $-\gamma$, $-\beta$, and $-\alpha$ rotating Z^P , Y' and X'' axes. Therefore, ω_{ec}^c can be given as

$$\omega_{ec}^c = (-\gamma/\Delta t)\vec{Z}^P + (-\beta/\Delta t)\vec{Y}' + (-\alpha/\Delta t)\vec{X}'' \tag{7}$$

Based on the rotation orders, the unit vectors of \vec{Y}' , \vec{X}'' can be derived as

$$\begin{aligned} \begin{bmatrix} \vec{X}' \\ \vec{Y}' \\ \vec{Z}' \end{bmatrix} &= \begin{bmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \vec{X}^c \\ \vec{Y}^c \\ \vec{Z}^c \end{bmatrix} \\ \begin{bmatrix} \vec{X}'' \\ \vec{Y}'' \\ \vec{Z}'' \end{bmatrix} &= \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix} \begin{bmatrix} \vec{X}' \\ \vec{Y}' \\ \vec{Z}' \end{bmatrix} \\ \vec{Y}' &= \sin \gamma \vec{X}^c + \cos \gamma \vec{Y}^c \\ \vec{X}'' &= \cos \beta \vec{X}' + \sin \beta \vec{Z}' = \cos \beta (\cos \gamma \vec{X}^c - \sin \gamma \vec{Y}^c) + \sin \beta \vec{Z}^c \end{aligned} \tag{8}$$

where γ , β , α are the Euler angles explained in Equation (1), namely Euler angles rotating Z^P , Y' and X'' respectively. With \vec{Y}' , \vec{X}'' obtained from Equation (8), Equation (7) can be rewritten as

$$\begin{aligned} \omega_{ec}^c &= (-\gamma/\Delta t)\vec{Z}^P + (-\beta/\Delta t)\vec{Y}' + (-\alpha/\Delta t)\vec{X}'' \\ &= \begin{bmatrix} -(\beta/\Delta t) \sin \gamma - (\alpha/\Delta t) \cos \beta \cos \gamma \\ -(\beta/\Delta t) \cos \gamma + (\alpha/\Delta t) \cos \beta \sin \gamma \\ -\gamma/\Delta t - (\alpha/\Delta t) \sin \beta \end{bmatrix} \end{aligned} \tag{9}$$

With Equations (5) to (9), the stereo VO can be updated based on velocity and angular rate.

3.2. Stereo VO error linearisation

Given the mechanisation process presented in Section 3.1, the errors of stereo VO mechanisation can be modelled by linearisation of Equation (5). The purpose of modelling the errors is to integrate different types of data. In this paper, the errors of position, velocity, attitude and angular velocity are estimated in the Kalman filter. The state vector is given as

$$X = [\delta r^e \ \delta v^c \ \epsilon \ \delta \omega_{ec}^c]^T \tag{10}$$

where δr^e is the position error in ECEF; δv^c is the velocity error in the camera frame; ϵ is the attitude error; $\delta \omega_{ec}^c$ is the angular rate error of the camera frame with respect to ECEF expressed in the camera frame.

The EKF procedure after linearisation can be summarised as (Welch and Bishop, 1995)

$$\begin{aligned} X_{k+1}^- &= \Phi_{k+1,k} X_k + w_k \\ P_{k+1}^- &= \Phi_{k+1,k} P_k \Phi_{k+1,k}^T + Q_k \\ K_{k+1} &= P_{k+1}^- H_{k+1}^T (H_{k+1} P_{k+1}^- H_{k+1}^T + R_{k+1})^{-1} \\ X_{k+1} &= X_{k+1}^- + K_{k+1} (Z_{k+1} - H_{k+1} X_{k+1}^-) \\ P_{k+1} &= (I - K_{k+1} H_{k+1}) P_{k+1}^- \end{aligned} \tag{11}$$

In Equation (11), X_{k+1}^- is the predicted state estimate with covariance matrix P_{k+1}^- . X_{k+1}^- is achieved based on the transition matrix $\Phi_{k+1,k}$ and the previously estimated state vector X_k with process noise w_k . The predicted covariance matrix P_{k+1}^- is achieved by uncertainty propagation of the previously estimated covariance matrix P_k and the process noise covariance matrix Q_k . The measurement update is represented as Z_{k+1} with covariance R_{k+1} and the corresponding design matrix is represented as H_{k+1} . Based on the optimal gain matrix K_{k+1} , the state vector and its covariance matrix can be updated. Note that in Equation (11) P has a different meaning to that in Equation (4).

Since the errors are estimated in the Kalman filter, Equation (5) needs to be linearised to achieve the dynamic information of the system errors, given as

$$\begin{bmatrix} \delta \dot{r}^e \\ \delta \dot{v}^c \\ \dot{\epsilon} \\ \delta \dot{\omega}_{ec}^c \end{bmatrix} = \begin{bmatrix} \Psi R_c^e v^c + R_c^e \delta v^c \\ w_v \\ R_c^e \delta \omega_{ec}^c \\ w_\omega \end{bmatrix} = \begin{bmatrix} -V^e \epsilon + R_c^e \delta v^c \\ w_v \\ R_c^e \delta \omega_{ec}^c \\ w_\omega \end{bmatrix} \tag{12}$$

where δ represents the error of each item; the dot on the left side of the equation represents the time derivative of each item; ϵ is the attitude error; Ψ is the skew-symmetric matrix of the attitude error ϵ ; $R_c^e v^c$ achieves the velocity in ECEF v^e ; V^e is the skew-symmetric matrix of v^e ; the velocity error δv^c and angular velocity error $\delta \omega_{ec}^c$ are modelled as random walk, w_v and w_ω are the driven noise of δv^c and $\delta \omega_{ec}^c$ respectively.

Based on Equation (12), the dynamic matrix of stereo VO can be formed as

$$F = \begin{bmatrix} 0 & R_c^e & -V^e & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & R_c^e \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{13}$$

The transition matrix $\Phi_{k+1,k}$ in Equation (11) can be simplified as

$$\Phi_{k+1,k} \approx I + F\Delta t = \begin{bmatrix} I & R_c^e \Delta t & -V^e \Delta t & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & R_c^e \Delta t \\ 0 & 0 & 0 & I \end{bmatrix} \tag{14}$$

Equation (13) shows the velocity error is not coupled with the angular rate error in the mechanisation. However, this is not the case for stereo VO. The velocity and angular rate should be correlated since the translation and rotation are correlated, based on Equations (3), (4), (6) and (9). To take full advantage of the information from PnP, the velocity and angular rate derived from Equations (6) and (9) are used to form the measurement update in the Kalman filter, which are the same as those used in mechanisation in Equation (5). Therefore, the measurement update would be $Z_{k+1} = 0$, which means that the measurement update does not have an impact on the Kalman filter estimation. However, the velocity and angular rate become coupled with each other due to the contribution of the measurement covariance. Specifically, the covariance matrix of the measurement update in Equation (11) can be written as

$$R_{k+1} = AR_{PnP}A^T$$

$$A = \begin{bmatrix} I/\Delta t & 0 \\ 0 & B \end{bmatrix}$$

$$B = \begin{bmatrix} -(1/\Delta t) \cos \beta \cos \gamma & -(1/\Delta t) \sin \gamma + (\alpha/\Delta t) \sin \beta \cos \gamma & -(\beta/\Delta t) \cos \gamma + (\alpha/\Delta t) \cos \beta \sin \gamma \\ (1/\Delta t) \cos \beta \sin \gamma & -(1/\Delta t) \cos \gamma - (\alpha/\Delta t) \sin \beta \sin \gamma & (\beta/\Delta t) \sin \gamma + (\alpha/\Delta t) \cos \beta \cos \gamma \\ -(1/\Delta t) \sin \beta & -(\alpha/\Delta t) \cos \beta & -(1/\Delta t) \end{bmatrix} \tag{15}$$

where R_{PnP} is the PnP covariance matrix achieved from Equation (4); A is the design matrix based on Equation (6) and linearisation of Equation (9); B is the linearisation of Equation (9).

For the process noise Q in Equation (11), the velocity and angular rate related process noise are set as large values to fully use the information from PnP. The design matrix in Equation (11) can be given as

$$H = \begin{bmatrix} 0 & I & 0 & 0 \\ 0 & 0 & 0 & I \end{bmatrix} \tag{16}$$

3.3. Ground vehicle velocity constraint

With the equations in Section 3.2, the stereo VO can be integrated with other types of measurements. When the stereo VO is applied on a ground vehicle, the velocity constraint can be applied to reduce the drift if no additional sensors are available. In most cases, there is only the forward velocity for a ground vehicle when it is not manoeuvring. Therefore, the rightward and downward velocity can be applied as constraints in land vehicle applications (Shin, 2001). The velocity constraints can correct the velocity, position and angular rate since they are coupled with each other. Furthermore, the attitude can be corrected as well since it is correlated with the angular rate. In this way, the performance of stereo VO can be improved.

The camera frame has to be aligned with the land vehicle frame with the X -axis of the camera frame pointing to the right of the ground vehicle and the Y -axis of the camera frame pointing downward of the ground vehicle. When the vehicle is not manoeuvring, the velocity constraint as measurement update can be given as (Shin, 2001; Godha and Cannon, 2007; Nouredin et al., 2012; Y. Liu et al., 2018)

$$Z_{NHC} = \begin{bmatrix} 0 - v_x^c \\ 0 - v_y^c \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} X \tag{17}$$

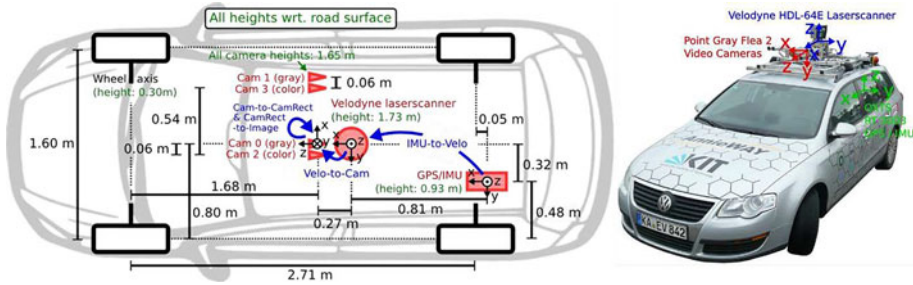


Figure 1. KITTI data collection system setup (Geiger et al., 2013).

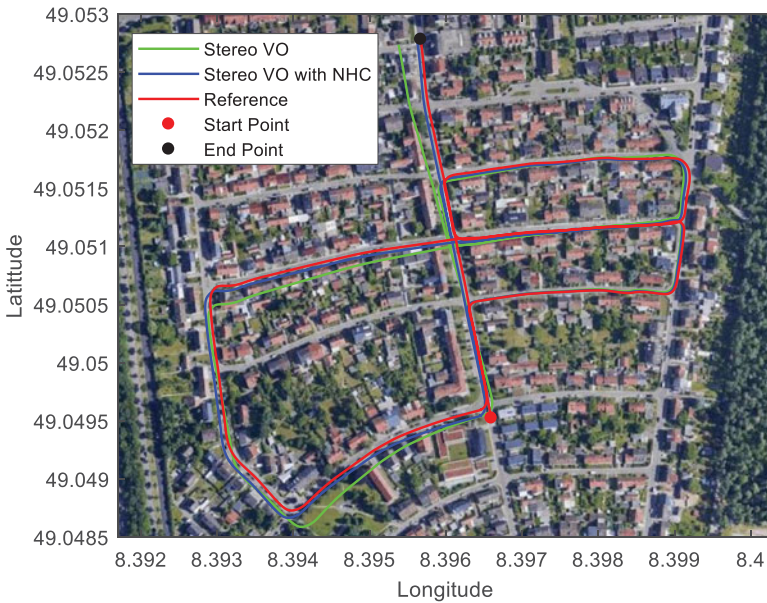


Figure 2. First dataset trajectories.

where Z_{NHC} is the velocity constraint, v_x^c and v_y^c are the x and y components of camera velocity; X is the state vector in Equation (10).

For ground vehicle navigation, horizontal positioning accuracy is much more important than height accuracy. The velocity in the X -axis and angular rate in Y -axis in the camera frame are coupled with each other and they have a greater impact on the horizontal positioning accuracy, compared with the velocity in the Y -axis. Therefore, Equation (17) can be further simplified as

$$Z_{NHC} = 0 - v_x^c \tag{18}$$

In this work, the angular rate of the Y -axis is applied to determine when to apply the constraint. If the magnitude of the angular rate is larger than the threshold, the constraint is not applied due to possible manoeuvres.

4. Field tests and results

Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) datasets (Geiger et al., 2013) are used to verify the proposed stereo VO with NHC for ground vehicle applications. Shown in Figure 1 is the KITTI test setup.

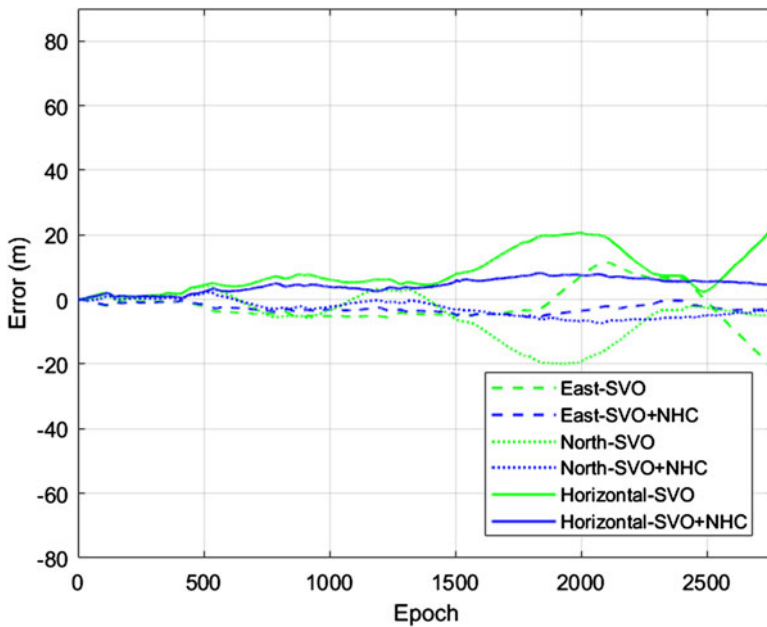


Figure 3. First dataset horizontal errors.

Table 1. RMS of stereo VO and stereo VO with NHC for the first dataset.

	Stereo VO		Stereo VO with NHC	
	RMS (m)	MAX (m)	RMS (m)	MAX (m)
East	6.08	21.27	3.11	5.70
North	8.33	19.93	3.81	7.38
Horizontal	10.32	21.89	4.92	8.16

In Figure 1, there are two grey (Cam 0 and Cam 1) and two colour cameras (Cam 2 and Cam 3) installed on the land vehicle roof. In this paper, only the grey stereo images collected by the grey cameras (Point Grey Flea 2 FL2-14S3M-C) are used. FL2-14S3M-C is a CCD (charge-coupled device) camera with a global shutter supporting 15 FPS (frames per second). The sensor format, pixel size and resolution are 1/2", 4.65 μm, and 1384 by 1032, respectively. The FPS used during the test is 10. The stereo images applied are calibrated, rectified and synchronised, which are all done by KITTI. The laser scanner shown in Figure 1 is not used in this work. The GNSS/IMU system OXTS RT 3003 is applied as the reference with an accuracy of 0.4 m CEP (circular error probable).

Three relatively long KITTI datasets collected in residential areas are applied in this work. To verify the improvement of the proposed method, the stereo VO with NHC results are compared with stand-alone stereo VO results. The reference used is the DGPS (differential GPS)/IMU output of RT 3003.

The first dataset contains 2,762 stereo images with a duration of about four and a half minutes. The total horizontal distance is approximately 2,205 m. The trajectories obtained by stereo VO and the proposed method together with the reference are plotted in Figure 2. The green, blue and red lines represent the stereo VO, stereo VO with NHC, and the reference (DGPS/IMU), respectively. It can be seen that although it is inevitable for both stereo VO and stereo VO with NHC to suffer drift, the accuracy of the proposed method with NHC is much improved. The trajectory of the proposed method is closer to the reference most time of the test.

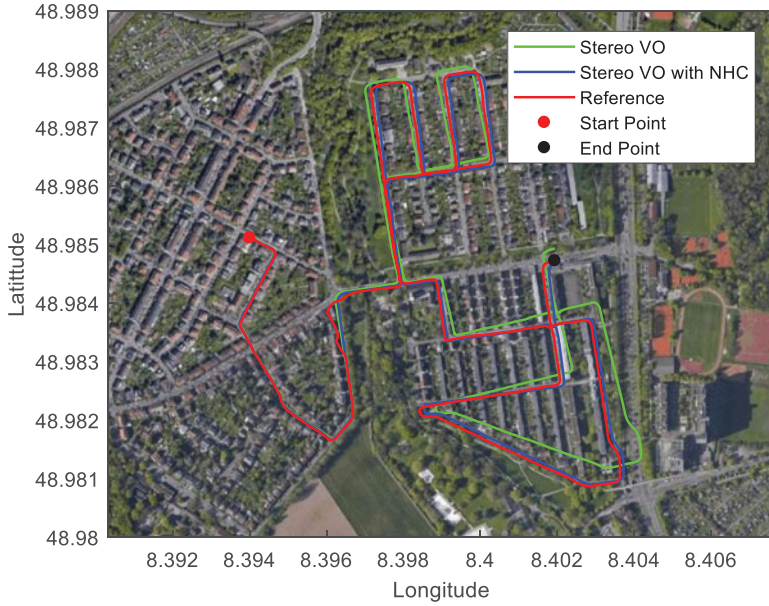


Figure 4. Second dataset trajectories.

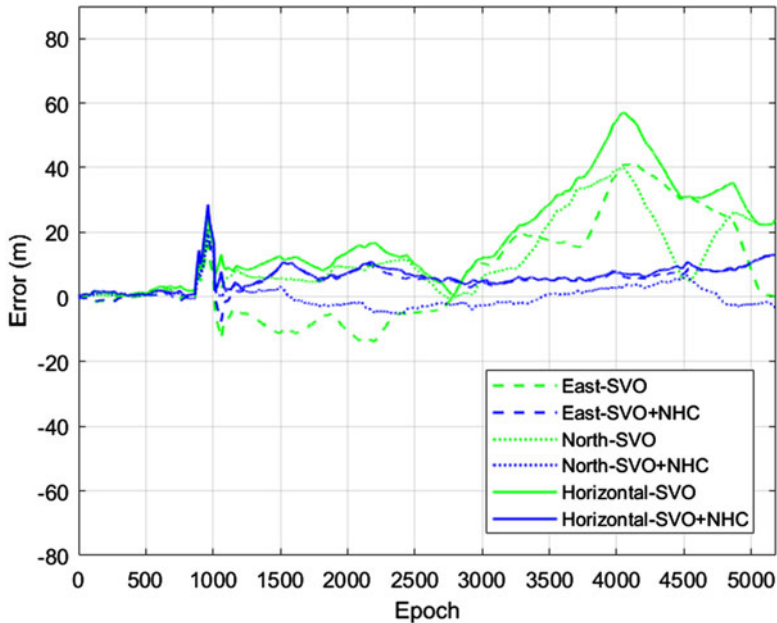


Figure 5. Second dataset horizontal errors.

To see the performance of the two different methods more clearly, the horizontal errors are plotted in Figure 3 and the root mean square (RMS) values of horizontal positioning error obtained by each method are summarised in Table 1. In Figure 3, the stereo VO and stereo VO with NHC errors are plotted in green and blue lines respectively. The horizontal errors are represented by solid lines while the east and north errors are plotted in two kinds of dash lines. Each epoch is 0.1 s since the FPS setting is 10. Generally, the errors in each direction for both methods increase with time. For stereo VO, the errors in east and north fluctuate during the test, which is caused by the manoeuvring. The maximum

Table 2. RMS of stereo VO and stereo VO with NHC for the second dataset.

	Stereo VO		Stereo VO with NHC	
	RMS (m)	MAX (m)	RMS (m)	MAX (m)
East	17.30	41.25	6.38	20.86
North	16.36	39.77	3.28	19.47
Horizontal	23.81	57.00	7.18	28.54

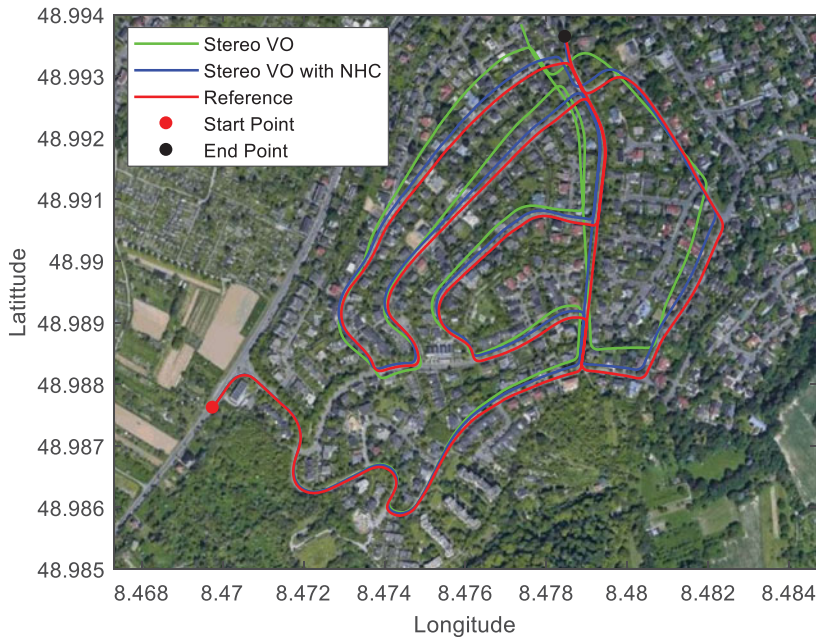


Figure 6. Third dataset trajectories.

horizontal error of stereo VO happens at the end of the trajectory. In comparison, the errors of stereo VO with NHC in each direction grow much more slowly with time and all the errors are under 10 m. The accuracy of the proposed method is improved in both east and north directions, compared with stereo VO. The horizontal RMS value of the proposed method is less than half of stereo VO. The maximum horizontal error is also reduced from 21.89 m to 8.16 m. From Figures 2 and 3, the stereo VO horizontal error is smaller than stereo VO with NHC in a short period only at around the 2500th epoch. After several manoeuvres, when the land vehicle just turns around, the stereo VO's position is not accurate due to the drift. However, it might move closer to the true path due to inaccurate heading. This is the reason that the stereo VO errors fluctuate and sometimes are slightly better than stereo VO with NHC.

The second dataset contains 5,177 stereo images with more manoeuvres during the test. The duration of this dataset is over eight and a half minutes and the horizontal distance is about 4,129 m. Similarly, in comparison with stereo VO, the trajectory of the proposed method with the application of the velocity constraint is much closer to the reference in Figure 4.

The horizontal errors of stereo VO and stereo VO with NHC are plotted in Figure 5 and summarised in Table 2. Stereo VO with NHC outperforms stereo VO, as shown in Figure 5. The abnormal peaks around the 900th epoch may be caused by the inaccurate positioning results of the reference DGPS/IMU system since the reference trajectory is not on the path during this short period. The maximum horizontal

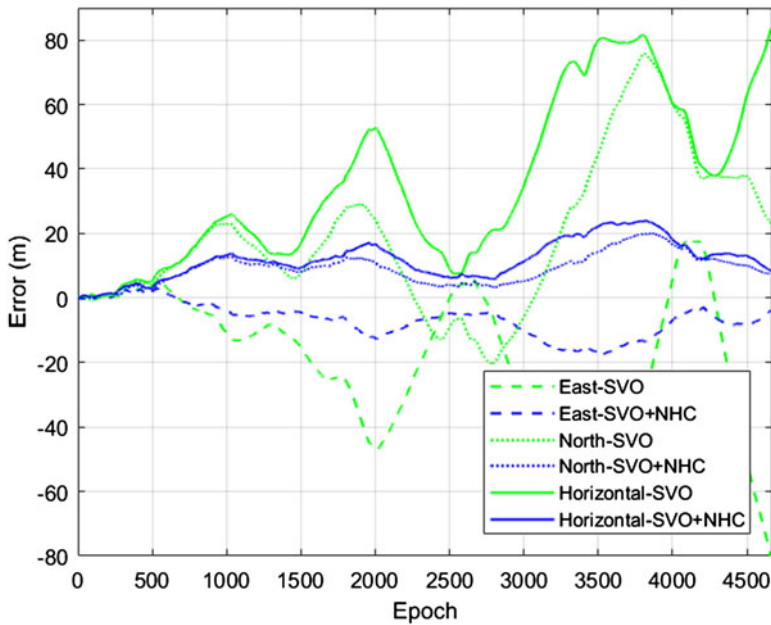


Figure 7. Third dataset horizontal errors.

error of stereo VO with NHC happens during this period. Other than this period, the errors of stereo VO with NHC in each direction grow slowly with time. Errors of stereo VO with NHC are under 10 m most of the time, while the error growth of stereo VO is much more evident, especially after the 3,500th epoch. Similar to the previous dataset, stereo VO errors are less than stereo VO with NHC during very short periods around the 2,700th epoch and 4,500th epoch. From Table 2, the RMS values of the proposed method are less than stereo VO in both east and north directions. The horizontal RMS value of the proposed method is only one-third of the stereo VO. The maximum horizontal error drops from 57 m to 28.54 m.

The third dataset contains 4,663 stereo images with the most manoeuvres. The horizontal distance of this data is about 5,060 m and the duration is more than seven and a half minutes. The trajectories are shown in Figure 6. Similarly, the drift of the proposed method is much reduced compared with stereo VO. To compare the accuracy of the two methods, the positioning errors with DGPS/IMU as the reference are plotted in Figure 7 and RMS values are listed in Table 3. Although the performance of the proposed method for this dataset is worse than the previous two datasets, the accuracy is still much improved with horizontal RMS almost one-third of stereo VO. Because there are more manoeuvres in this test, the horizontal errors of stereo VO fluctuate more than the previous datasets. The general error trends of stereo VO with NHC are similar to the stereo VO, but the magnitudes are much less. The velocity constraint effectively limits the error growth with a horizontal error smaller than 20 m most of the time while the stereo VO errors increase significantly, especially during the second half of the test. With the application of velocity constraint, the horizontal error RMS is reduced from 43.17 m to 13.24 m while the maximum horizontal error drops from 83.48 m to 23.88 m.

To summarise the results in this section, the relative horizontal position errors (the ratio of horizontal error RMS to the horizontal distance) of both methods are listed in Table 4. The relative horizontal position errors of the proposed method are much reduced compared with stereo VO. The relative horizontal position errors of stereo VO with NHC are all less than half of the stereo VO. With the velocity constraint applied, the relative horizontal position errors are all reduced to less than 0.3%.

Table 3. RMS of stereo VO and stereo VO with NHC for the third dataset.

	Stereo VO		Stereo VO with NHC	
	RMS (m)	MAX (m)	RMS (m)	MAX (m)
East	31.13	80.19	8.44	17.53
North	29.91	75.77	10.20	19.99
Horizontal	43.17	83.48	13.24	23.88

Table 4. Relative horizontal position errors for the datasets.

Dataset	Distance	Stereo VO	Stereo VO with NHC
1	2205 m	0.47%	0.22%
2	4129 m	0.58%	0.17%
3	5060 m	0.85%	0.26%

5. Conclusion

This paper proposes a velocity constrained stereo VO method to improve the horizontal positioning accuracy for ground vehicle navigation. The method to apply the stereo VO as the core component in multi-sensor integration is presented. The drift of stand-alone stereo VO is effectively reduced by introducing the velocity constraint of the ground vehicle. The test results with the KITTI datasets show that the horizontal errors of the proposed method are much less than the stereo VO. By comparison with stereo VO, both the overall horizontal error RMS and maximum horizontal errors are reduced to half of the stand-alone stereo VO. The relative horizontal position errors of the proposed method are reduced to less than 0.3% for all the datasets.

References

- Balazadegan Sarvrood, Y.** (2016) Multi-Sensor Map Matching Techniques for Autonomous Land Vehicle Navigation. University of Calgary, Calgary, Canada.
- Bay, H., Tuytelaars, T. and Van Gool, L.** (2006). Surf: Speeded up robust features. In: Leonardis, A., Bischof, H and Pinz, A (eds.). *Computer Vision—ECCV 2006*, Berlin Heidelberg: Springer, pp. 404–417.
- Bloesch, M., Omari, S., Hutter, M., Siegwart, R.** (2015). Robust Visual Inertial Odometry Using a Direct EKF-Based Approach. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 298–304.
- Bloesch, M., Burri, M., Omari, S., Hutter, M., Siegwart, R.** (2017). Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback. *The International Journal of Robotics Research*, **36**(10), 1053–1072.
- Campos, C., Elvira, R., Rodríguez, J.J.G., Montiel, J.M., Tardós, J.D.** (2020). ORB-SLAM3: An accurate open-source library for visual, visual-inertial and multi-map SLAM. *arXiv preprint arXiv:2007.11898*.
- Cvišić, I. and Petrović, I.** (2015). Stereo Odometry Based on Careful Feature Selection and Tracking. In: *2015 European Conference on Mobile Robots (ECMR)*. IEEE, pp. 1–6.
- Fischler, M. A. and Bolles, R. C.** (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, **24**(6), 381–395.
- Forster, C., Zhang, Z., Gassner, M., Werlberger, M., Scaramuzza, D.** (2016). SVO: Semidirect visual odometry for monocular and multicamera systems. *IEEE Transactions on Robotics*, **33**(2), 249–265.
- Gálvez-López, D. and Tardos, J. D.** (2012). Bags of binary words for fast place recognition in image sequences. *IEEE Transactions on Robotics*, **28**(5), 1188–1197.
- Gao, X.S., Hou, X.R., Tang, J., Cheng, H.F.** (2003). Complete solution classification for the perspective-three-point problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25**(8), 930–943.
- Geiger, A., Lenz, P., Stiller, C., Urtasun, R.** (2013). Vision meets robotics: The KITTI dataset. *The International Journal of Robotics Research*, **32**(11), 1231–1237.
- Godha, S. and Cannon, M. E.** (2007). GPS/MEMS INS integrated system for navigation in urban areas. *GPS Solutions*, **11**(3), 193–203.

- Groves, P. D. and Jiang, Z. (2013). Height aiding, C/N 0 weighting and consistency checking for GNSS NLOS and multipath mitigation in urban areas. *The Journal of Navigation*, **66**(5), 653–669.
- Hartley, R. and Zisserman, A. (2001). *Multiple View Geometry in Computer Vision*. Cambridge: Cambridge University Press.
- Hesch, J. A. and Roumeliotis, S. I. (2011). A Direct Least-Squares (DLS) Method for PnP. In: *2011 International Conference on Computer Vision*. IEEE, pp. 383–390.
- Jekeli, C. (2012). *Inertial Navigation Systems with Geodetic Applications*. Berlin: Walter de Gruyter.
- Lepetit, V., Moreno-Noguer, F. and Fua, P. (2009). EPnP: Efficient perspective-n-point camera pose estimation. *International Journal of Computer Vision*, **81**(2), 155–166.
- Leutenegger, S. (2015). Keyframe-based visual-inertial odometry using nonlinear optimization. *The International Journal of Robotics Research*, **34**(3), 314–334.
- Liu, F. (2018). Tightly Coupled Integration of GNSS/INS/Stereo Vision/Map Matching System for Land Vehicle Navigation. University of Calgary, Calgary, Canada.
- Liu, F., Balazadegan Sarvrood, Y. and Gao, Y. (2018a). Tight integration of INS/stereo VO/digital map for land vehicle navigation. *Photogrammetric Engineering & Remote Sensing*, **84**(1), 15–23.
- Liu, F., Balazadegan Sarvrood, Y. and Gao, Y. (2018b). Implementation and analysis of tightly integrated INS/stereo VO for land vehicle navigation. *The Journal of Navigation*, **71**(1), 83–99.
- Liu, Y., Gu, Y., Li, J., Zhang, X. (2017). Robust stereo visual odometry using improved RANSAC-based methods for mobile robot localisation. *Sensors*, **17**(10), 2339.
- Liu, Y., Liu, F., Gao, Y., Zhao, L. (2018). Implementation and analysis of tightly coupled global navigation satellite system precise point positioning/inertial navigation system (GNSS PPP/INS) with insufficient satellites for land vehicle navigation. *Sensors*, **18**(12), 4305.
- Menze, M. and Geiger, A. (2015). Object Scene Flow for Autonomous Vehicles. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3061–3070.
- Nistér, D. (2005). Preemptive RANSAC for live structure and motion estimation. *Machine Vision and Applications*, **16**(5), 321–329.
- Nistér, D., Naroditsky, O. and Bergen, J. (2004). Visual Odometry. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004*. CVPR 2004. IEEE, pp. 1–652.
- Noureddin, A., Karamat, T. B. and Georgy, J. (2012). *Fundamentals of Inertial Navigation, Satellite-Based Positioning and Their Integration*. Berlin Heidelberg: Springer Science & Business Media.
- Qin, T., Li, P. and Shen, S. (2018). VINS-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, **34**(4), 1004–1020.
- Qin, T., Pan, J., Cao, S., Shen, S. (2019). A general optimization-based framework for local odometry estimation with multiple sensors. *arXiv preprint arXiv:1901.03638*.
- Rosten, E. and Drummond, T. (2006). Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H. and Pinz, A. (eds.). *Computer Vision—ECCV 2006*, Berlin Heidelberg: Springer, pp. 430–443.
- Scaramuzza, D. (2011). Performance evaluation of 1-point-RANSAC visual odometry. *Journal of Field Robotics*, **28**(5), 792–811.
- Scaramuzza, D. and Fraundorfer, F. (2011). Visual odometry [tutorial]. *IEEE Robotics & Automation Magazine*, **18**(4), 80–92.
- Shi, J. and Tomasi, C. (1993). *Good Features to Track*. New York: Cornell University.
- Shin, E.-H. (2001). Accuracy Improvement of Low Cost INS/GPS for Land Applications. UCGE Report. University of Calgary.
- Strasdat, H., Montiel, J. M. M. and Davison, A. J. (2012). Visual SLAM: Why filter? *Image and Vision Computing*, **30**(2), 65–77.
- Tomasi, C. and Kanade, T. (1992). Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, **9**(2), 137–154.
- Von Stumberg, L., Usenko, V. and Cremers, D. (2018). Direct Sparse Visual-Inertial Odometry Using Dynamic Marginalization. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 2510–2517.
- Wang, R., Schworer, M. and Cremers, D. (2017). Stereo DSO: Large-Scale Direct Sparse Visual Odometry with Stereo Cameras. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3903–3911.
- Welch, G. and Bishop, G. (1995). *An Introduction to the Kalman Filter*. Chapel Hill, North Carolina: Department of Computer Science, University of North Carolina at Chapel Hill.
- Zhang, J., Kaess, M. and Singh, S. (2014). Real-Time Depth Enhanced Monocular Odometry. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, pp. 4973–4980. doi: 10.1109/IROS.2014.6943269.