Heading Reference-Assisted Pose Estimation for Ground Vehicles

Article *in* IEEE Transactions on Automation Science and Engineering · April 2018

DOI: 10.1109/TASE.2018.2828078

CITATIONS
7

READS
116

4 authors, including:

Han Wang
University of Virginia
45 PUBLICATIONS **325** CITATIONS
SEE PROFILE

Rui Jiang
National University of Singapore
21 PUBLICATIONS **129** CITATIONS
SEE PROFILE

Handuo Zhang
Nanyang Technological University
11 PUBLICATIONS **136** CITATIONS
SEE PROFILE

Heading Reference Assisted Pose Estimation for Ground Vehicles

Han Wang[1], *Senior Member, IEEE*, Rui Jiang[1,2], Handuo Zhang[1], Shuzhi Sam Ge[2], *Fellow, IEEE*

***Abstract*—In this paper, heading reference assisted pose estimation has been proposed to compensate inherent drift of Visual Odometry (VO) on ground vehicles, where estimation error is prone to grow while the vehicle is making turns or in environments with poor features. By introducing a particular orientation as "heading reference", a pose estimation framework has been presented to incorporate measurements from heading reference sensors into VO. A graph formulation is then proposed to represent the pose estimation problem under the commonly-used graph optimization model. Simulations and experiments on KITTI dataset and our self-collected sequences have been conducted to verify the accuracy and robustness of the proposed scheme. KITTI sequences and manually-generated heading measurement with Gaussian noises are used in simulation, where rotational drift error is observed to be bounded. Compared to pure VO, the proposed approach greatly reduces average translational localization error from 153.85 m to 24.29 m and 23.80 m in self-collected stereo visual sequences with traveling distance over 4.5 km at processing rates 19.7 Hz and 11.1 Hz, for the loosely-coupled and tightly-coupled model, respectively.**

***Note to Practitioners*— When Global Positioning System (GPS) is not available or reliable, Visual Odometry (VO) on ground vehicles is an efficient tool for estimating the pose, which involves translation and rotation. However, VO inherently suffers from drifting issue due to constant iterations. By adding a low-cost heading reference sensor, this work first introduces graph optimization formulation of pose estimation, then presents a pose estimation framework which incorporates heading measurements to VO such that long-term translation and rotation estimation errors can be greatly reduced in real-time computation. As a supplementary to VO, performance may still deteriorate in environments with poor illumination conditions and high-frequency movements. The proposed approach may be further improved by fusing heading measurements from more sensors, or being used to build a heading reference assisted Simultaneous Localization and Mapping (SLAM) system on any off-the-shelf SLAM framework.**

***Index Terms*—visual odometry, heading reference, robot navigation, pose estimation, graph optimization**

I. Introduction

VISUAL Odometry (VO), as a promising pose estimation approach, has been extensively used in robots and autonomous ground vehicles [1], [2]. Although prominent achievement has been made in test datasets [3], it is still challenging to implement VO in environments with poor illumination conditions, insufficient features, and dynamic scenes.

Translation and rotation error, which both cause error accumulation in VO, are closely coupled. A small drift in rotation may result in abysmal VO performance. This paper works on assisting VO by improving rotation estimation with an orientation reference. Although Visual-Inertial Odometry (VIO) has performed satisfactorily, components in VIO are both based on dead-reckoning which is inherently sensitive to drifting issue. Thus, it is necessary to define an orientation in the global frame to provide an absolute reference for drift suppression. The recent development of micro-electromechanical technologies enables compact, precise and cost-efficient Attitude and Heading Reference Systems (AHRS), which provides accurate and timely orientation estimation. Differentiated from VIO, in this paper, a pose estimation framework is proposed, where an absolute heading reference is used to assist VO such that robust and accurate vehicle localization can be attained.

The contributions of this work are three-fold:

1) We propose to use the absolute heading in VO to suppress rotation drift. By utilizing the coupling characteristics between rotation and translation in VO, the proposed heading measurement can be cost-effectively used for benefiting both rotation and translation estimation in ground vehicles.
2) A sensor fusion framework has been proposed to incorporate heading measurements into visual odometry. In the framework, a heading reference is abstracted as a vertex in graph model, based on which, the problem is formulated as graph optimization such that off-the-shelf back-end libraries can be utilized effortlessly.
3) To demonstrate the effectiveness of heading reference assisted approach, extensive simulations and experiments have been conducted based on KITTI dataset and self-collected data. The results are compared and discussed between pure VO and the proposed approach.

The remaining part of the paper is organized as follows: After presenting related work in Section II, we introduce principles of VO and heading estimation, then the graph optimization problem is formulated at the end of Section III. Section IV provides two levels of abstraction models involving heading reference. In Section V, a general framework of pose estimation with heading reference is proposed. Section VI and VII detail the implementation of the proposed approach and demonstrate the experimental results. Finally, VIII concludes the paper and discusses on possible future work.

II. Related Work

The existing solution on enhancing VO performance falls on i) improving VO components including feature detection,

The first two authors contributed equally to this work.

[1]School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798. hw@ntu.edu.sg, csjiangrui@gmail.com

[2]Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117583. rui_jiang@u.nus.edu, samge@nus.edu.sg

matching, outlier removal or pose optimization; and ii) seeking assistance from other approaches or databases [4] such as lidar [5], Global Positioning System (GPS) [6], digital maps [7], [8], Inertial Navigation System (INS) and many others [9]–[20]. Benefiting from its self-contained property, many Visual-Inertial Odometry (VIO) schemes have been proposed to reduce drift in VO. Loosely-coupled methods [10], [13] fuse data at a higher level, where data from Inertial Measurement Unit (IMU) and VO are fused after being obtained respectively; Tightly-coupled methods, which consider not only pose but features as state variables in estimation, generally achieve higher precision but also suffer from heavier computational cost. There are two main streams in tightly-coupled VIO: On the one hand, a filter-based method is proposed to estimate egomotion, camera extrinsic parameters as well as the additive IMU biases in [9]. On the other hand, by optimization-based methods, pose estimation can be formulated as a non-linear least square optimization problem which aims to minimize a cost function containing inertial error terms and reprojection error simultaneously. Authors in [14] have proposed an integration framework where the concept of keyframes and marginalization are introduced to ensure real-time operation. In [16], a preintegration scheme of inertial measurement between keyframes has been proposed, where fused measurement model and error propagation expression have been derived such that the optimization could be achieved directly on-manifold. Authors in [21] have addressed the initilization and calibration problems on the fly for monucular VIO. Unfortunately, existing methods still suffer from drift issue, which motivates us to eliminate rotation drift by introducing an absolute heading in pose estimation.

As for deployment of orientations as supplementary information, gravity has been used as a vertical reference for vision and inertial sensor cooperation in Structure from Motion (SfM) methods, where image horizon line can be determined [11]. Authors in [22] have further proposed an egomotion estimation approach, where fewer point correspondences for relative motion estimation are needed by utilizing the gravity direction.To the best of authors' knowledge, this paper serves as the first attempt to discuss heading reference assisted pose estimation problem for ground vehicles.

III. Preliminaries

In this section, principles of stereo VO, heading estimation and graph optimization are introduced in preparation for heading reference assisted pose estimation.

A. Visual Odometry

As an integrated sensor, VO can be modelled as [23]

$$\mathtt{C}_{k+1} = \mathtt{C}_k \mathtt{T}_{k,k+1} \tag{1}$$

where $\mathtt{C}_k = \left[\mathtt{R}_k | \mathbf{t}_k\right] \in \mathrm{SE}(3)$ denotes camera pose matrix at time instant k; $\mathtt{T}_{k,k+1} \in \mathrm{SE}(3)$, which is the measurement, denotes the transformation matrix between pose from k to $k+1$; $\mathtt{R}_k \in \mathrm{SO}(3)$ and $\mathbf{t}_k \in \mathbb{R}^3$ are rotation matrix and translation vector at k, respectively.

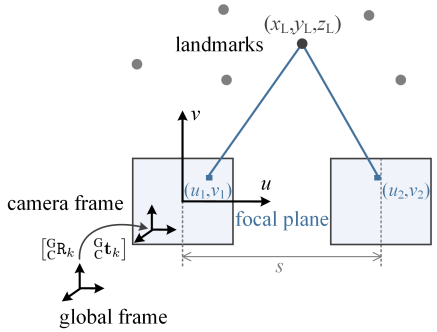


Fig. 1. Stereo 3D-2D projection from Euclidean space to focal planes.

Taking VO as an integrated module may be insufficient in scenarios where feature-level optimization is necessary. With an image pair from a stereo camera system, the coordinates of 3D feature points (or so-called "landmarks") can be recovered. The 3D-3D correspondences could be utilized to calculate camera motion. Suppose the landmark is represented as $\mathbf{x}_L = [x_L, y_L, z_L]^\top$ in global frame, the ideal measurement equation that projects landmarks to homogeneous image coordinates on focal plane are given as follows [24] and illustrated in Fig. 1.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & c_u \\ 0 & f & c_v \\ 0 & 0 & 1 \end{bmatrix} \left(\begin{bmatrix} {}^{G}_{C}\mathtt{R}_k & {}^{G}_{C}\mathbf{t}_k \end{bmatrix} \begin{bmatrix} x_L \\ y_L \\ z_L \\ 1 \end{bmatrix} - \begin{bmatrix} s \\ 0 \\ 0 \end{bmatrix} \right) \tag{2}$$

with focal length f, projection center $[c_u, c_v]^\top$, rotation matrix and translation vector from global frame to camera frame ${}^{G}_{C}\mathtt{R}_k$ and ${}^{G}_{C}\mathbf{t}_k$, shift $s = 0$ for left images and s equaling to baseline for right images.

B. Heading Estimation Using Vector Measurements

Herein, the rotation from camera frame to East-North-Up (ENU) frame is taken as heading measurements for ground vehicles. Rotation between two frames can be estimated by measuring ambient vector field in different frames. Without loss of generality, let us suppose origins of all frames coincide with each other. The static ambient vector field $\mathbf{v} \in \mathbb{R}^3$ is written as ${}_{G}\mathbf{v}$ in global frame. By measuring $\mathbf{v}$ in body frame as ${}_{B}\mathbf{v}$, the heading of the body frame w.r.t. the global frame can be represented as a rotation matrix ${}^{B}_{G}\mathtt{R}$ which satisfies ${}_{G}\mathbf{v} = {}^{B}_{G}\mathtt{R}\,{}_{B}\mathbf{v}$. The angle and unified axis of the rotation are derived as

$$\theta = \arccos\left(\frac{{}_{B}\mathbf{v} \cdot {}_{G}\mathbf{v}}{\| {}_{B}\mathbf{v} \| \| {}_{G}\mathbf{v} \|} \right) \tag{3}$$

$$\mathbf{a} = \frac{{}_{B}\mathbf{v} \times {}_{G}\mathbf{v}}{\| {}_{B}\mathbf{v} \times {}_{G}\mathbf{v} \|} \tag{4}$$

The rotation matrix can be obtained using exponential map as

$${}^{B}_{G}\mathtt{R} = \exp\left([\mathbf{a}]_\times \theta\right) = \mathbf{I}_3 + \sin\theta [\mathbf{a}]_\times + (1 - \cos\theta)[\mathbf{a}]^2_\times \tag{5}$$

where $\mathbf{I}_3$ is the identity matrix of size 3; $[\mathbf{a}]_\times$ denotes the skew-symmetric matrix (or "cross-product matrix") for vector $\mathbf{a}$. Unfortunately, by noticing the norm constraint $\| {}_{B}\mathbf{v} = {}_{G}\mathbf{v} \|$, equation system ${}_{G}\mathbf{v} = {}^{B}_{G}\mathtt{R}\,{}_{B}\mathbf{v}$ is underdetermined. In other words, at least two vector measurements are required for the same time to determine a unique ${}^{B}_{G}\mathtt{R}$. A general method to

obtain the solution would be minimizing the loss function $J({}^{\mathrm{B}}_{\mathrm{G}}\mathrm{R}) = \frac{1}{2}\sum w_k \| {}_{\mathrm{G}}\mathbf{v} - {}^{\mathrm{B}}_{\mathrm{G}}\mathrm{R}\,{}_{\mathrm{B}}\mathbf{v}\|^2$ from multiple sensors where w_k denote weight for sensor k (known as Wahba's problem) [25].

Since the motion range of most applications is limited in a small-scale space where Earth's magnetic field and gravitational field can be deemed as constant, magnetometer and accelerometer are typical components in heading reference sensors. In practice, a low-pass filter is required to extract gravity from accelerometer values such that the influence of motion could be filtered out. To achieve high-frequency data processing at up to 150 Hz, QUEST (QUaternion ESTimator) [26] has been extensively used, which gives direct quaternion sub-optimal estimation as

$$\mathbf{q} = \frac{1}{\sqrt{1+\mathbf{p}^\top\mathbf{p}}}\begin{bmatrix}\mathbf{p}\\ 1\end{bmatrix} \tag{6}$$

where $\mathbf{p} = [(\sum w_k + \sigma)\mathbf{I} - \mathbf{S}]^{-1}\mathbf{Z}$, $\sigma = \mathrm{tr}(\mathbf{B})$, $\mathbf{S} = \mathbf{B} + \mathbf{B}^\top$, $\mathbf{B} = \sum w_k \left({}_{\mathrm{G}}\mathbf{v}\,{}_{\mathrm{B}}\mathbf{v}^\top\right)$, $\mathbf{Z} = [B_{23}-B_{32}, B_{31}-B_{13}, B_{12}-B_{21}]^\top$.

C. Graph Optimization on Manifold

Graph is a frequently-used model to represent measurements and states. Let $\mathbf{x}_i$ describe the state of vertex i, $\mathbf{h}_{ij}(\mathbf{x}_i, \mathbf{x}_j)$ be the ideal measurement equation, and $\mathbf{z}_{ij}$ denote actual measurement between vertex i and vertex j. With the existence of measurement error, we can always define an error function as $\mathbf{e}_{ij}(\mathbf{z}_{ij}, \mathbf{x}_i, \mathbf{x}_j) = \mathbf{z}_{ij} \ominus \mathbf{h}_{ij}(\mathbf{x}_i, \mathbf{x}_j)$ where $\ominus$ is an operator measuring difference between ideal and actual measurements. To simplify the notation, we let $\mathbf{x} = [\mathbf{x}_1^\top, \cdots, \mathbf{x}_n^\top]^\top$ be the parameter vector indicating n vertices and $\mathbf{x}_{ij} = [\mathbf{x}_i^\top, \mathbf{x}_j^\top]^\top$. The graph optimization problem aims to obtain the optimal $\mathbf{x}$ as

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \sum_{<i,j>\in\mathcal{C}} \underbrace{\mathbf{e}_{ij}(\mathbf{z}_{ij}, \mathbf{x}_{ij})^\top \mathbf{\Omega}_{ij} \mathbf{e}_{ij}(\mathbf{z}_{ij}, \mathbf{x}_{ij})}_{\mathbf{F}_{ij}} \tag{7}$$

where $\mathcal{C}$ denotes the index set containing measurements, $\mathbf{\Omega}_{ij}$ represent the information matrix of measurement $\mathbf{z}_{ij}$.

As it is difficult to find an analytical solution to (7) due to nonlinearity, iterations are typically required until a numerical solution is obtained. In Euclidean space, given an initial guess $\mathring{\mathbf{x}}_{ij}$, the error function around $\mathring{\mathbf{x}}_{ij}$ can be approximated as

$$\mathbf{e}_{ij}(\mathring{\mathbf{x}}_{ij} + \Delta\mathbf{x}_{ij}) \simeq \mathbf{e}_{ij}(\mathring{\mathbf{x}}_{ij}) + \mathbf{J}_{ij}\Delta\mathbf{x}_{ij} \tag{8}$$

where $\mathbf{J}_{ij}$ is the Jacobian of $\mathbf{e}_{ij}$ at $\mathring{\mathbf{x}}_{ij}$. The term $\mathbf{z}_{ij}$ in cost function is omitted as we are optimizing $\mathbf{x}_{ij}$ instead of the measurement. Taking (8) into $\mathbf{F}_{ij}$ leads to

$$\begin{aligned}\mathbf{F}_{ij}(\mathring{\mathbf{x}}_{ij} + \Delta\mathbf{x}_{ij}) &= \underbrace{\mathbf{e}_{ij}^\top(\mathring{\mathbf{x}}_{ij})\mathbf{\Omega}_{ij}\mathbf{e}_{ij}(\mathring{\mathbf{x}}_{ij})}_{c_{ij}} + \\ 2\underbrace{\mathbf{e}_{ij}^\top(\mathring{\mathbf{x}}_{ij})\mathbf{\Omega}_{ij}\mathbf{J}_{ij}}_{\mathbf{b}_{ij}^\top}\Delta\mathbf{x}_{ij} &+ \Delta\mathbf{x}_{ij}^\top\underbrace{\mathbf{J}_{ij}^\top\mathbf{\Omega}_{ij}\mathbf{J}_{ij}}_{\mathbf{H}_{ij}}\Delta\mathbf{x}_{ij} \qquad (9)\\ &= c_{ij} + 2\mathbf{b}_{ij}^\top\Delta\mathbf{x}_{ij} + \Delta\mathbf{x}_{ij}^\top\mathbf{H}_{ij}\Delta\mathbf{x}_{ij} \qquad (10)\end{aligned}$$

The summed cost function can be written as

$$\begin{aligned}\mathbf{F}(\mathring{\mathbf{x}} + \Delta\mathbf{x}) &= \sum_{<i,j>\in\mathcal{C}} \mathbf{F}_{ij}(\mathring{\mathbf{x}}_{ij} + \Delta\mathbf{x}_{ij}) \qquad (11)\\ &\simeq c + 2\mathbf{b}^\top\Delta\mathbf{x} + \Delta\mathbf{x}^\top\mathbf{H}\Delta\mathbf{x} \qquad (12)\end{aligned}$$

where $c = \sum c_{ij}$, $\mathbf{b} = \sum \mathbf{b}_{ij}$, $\mathbf{H} = \sum \mathbf{H}_{ij}$. Taking the derivative of $\mathbf{F}(\mathring{\mathbf{x}} + \Delta\mathbf{x})$ w.r.t. $\Delta\mathbf{x}$ leads to a sparse linear system $\mathbf{H}\Delta\mathbf{x}^* = -\mathbf{b}$, from which $\Delta\mathbf{x}^*$ could be obtained as the optimal increment to initial guess of the whole state $[\Delta\mathbf{x}_1^{*\top}, \cdots, \Delta\mathbf{x}_M^{*\top}]^\top$ with M vertices. Then the linearized optimal solution is obtained as

$$\mathbf{x}^* = \mathring{\mathbf{x}} + \Delta\mathbf{x}^* \tag{13}$$

When states are parameterized in non-Euclidean space, a similar optimization procedure can be carried out on a manifold, which is regarded as Euclidean space locally but not globally. To represent optimization problem on manifold, an operator $\boxplus$ needs to be defined to replace the simple addition such that the state after a small perturbation $\mathbf{x}\boxplus\Delta\mathbf{x}$ is still on the manifold. The error function on manifold can be modified as

$$\mathbf{e}_{ij}(\mathring{\mathbf{x}}_{ij} \boxplus \Delta\mathbf{x}_{ij}) \simeq \mathbf{e}_{ij}(\mathring{\mathbf{x}}_{ij}) + \tilde{\mathbf{J}}_{ij}\Delta\mathbf{x}_{ij} \tag{14}$$

where $\tilde{\mathbf{J}}_{ij} = \frac{\partial \mathbf{e}_{ij}(\mathring{\mathbf{x}}_{ij}\boxplus\Delta\mathbf{x}_{ij})}{\partial\Delta\mathbf{x}_{ij}}|_{\Delta\mathbf{x}_{ij}=0}$. In a VO system, to maintain rotational constraints, the small rotation increment is expressed minimally using Euler angles; To avoid singularity, the states are represented in over-parameterized space $\mathrm{SE}(3)$. Accordingly, given an initial guess $\mathring{\mathbf{x}}$, the states can be updated from $\mathbf{x}^* = \mathring{\mathbf{x}} \boxplus \Delta\mathbf{x}^*$ after obtaining the optimal increment $\Delta\mathbf{x}^*$.

To facilitate fast implementation, several computing packages and libraries have been developed to solve the graph optimization problem [27] [28] [29].

D. Problem Formulation

We consider the visual pose estimation problem for ground vehicles by graph optimization, where vehicle translation and rotation are unknown parameters to be solved. In other words, we aim at optimizing camera pose $\mathtt{C}_k$ at time k, given measurement set $\{\mathbf{z}_{ij}\}, \forall <i,j> \in \mathcal{C}_k$ where $\mathcal{C}_k$ denotes the index set containing all available measurements until time k.

IV. Abstraction Model of Measurement with Heading Reference

Redesign of graph model is required when heading measurements are brought in. In this section, we propose two types of graph abstraction models which both incorporate heading reference and heading measurements into the visual pose estimation framework such that the existing graph models are generalized. Loosely-coupled model deems VO a black box, whose input and output are stereo images and transformation matrix, respectively. At a lower abstract level, the tightly-coupled model takes 2D landmarks on image planes as observations, which enable us to consider constraints between landmarks.

A. Loosely-Coupled Model

The loosely-coupled model is illustrated in Fig. 2(a). Let $\mathbf{x} = [\mathbf{x}_0^\top, \mathbf{x}_1^\top, \cdots, \mathbf{x}_N^\top]^\top$ be the state, where $\mathbf{x}_0 \in \mathbb{R}^{3\dagger}$ denotes

†We show vector dimension with a slight abuse of notation. Actually, the range of elements in unit quaternion does not cover $\mathbb{R}$. The same notation goes to $\mathbf{x}_1, \cdots, \mathbf{x}_N$, $\mathbf{z}_{0k}$, $\mathbf{z}_{k,k+1}$ and overloading functions `v2t`, `t2v`.

heading reference in three-dimensional unit quaternion vector and $\mathbf{x}_1, \cdots, \mathbf{x}_N \in \mathbb{R}^6$ are camera poses vectors with translation and rotation components. Let $\mathbf{z}_{ij}$ be the measurement, where $\mathbf{z}_{0k} \in \mathbb{R}^3$ represents rotation from ENU frame to camera frame in unit quaternion vector, and $\mathbf{z}_{k,k+1} \in \mathbb{R}^6$ denotes transformation vector between consecutive poses at time k and $k+1$ measured by VO.

By defining an overloading function $\mathtt{v2t} : \mathbb{R}^3 \rightarrow \mathrm{SO}(3); \mathbb{R}^6 \rightarrow \mathrm{SE}(3)$ which converts vector representation to rotation or transformation matrix, we write $\mathtt{R}_0 = \mathtt{v2t}(\mathbf{x}_0)$, $\mathtt{C}_k = \mathtt{v2t}(\mathbf{x}_k)$, $\mathtt{R}_{0k} = \mathtt{v2t}(\mathbf{z}_{0k})$, and $\mathtt{T}_{k,k+1} = \mathtt{v2t}(\mathbf{z}_{k,k+1})$. The ideal measurement equation for VO and heading are:

$$\mathbf{h}^{\mathrm{VO}}(\mathtt{C}_k, \mathtt{C}_{k+1}) = \mathtt{C}_k^{-1}\mathtt{C}_{k+1} \tag{15}$$

$$\mathbf{h}^{\mathrm{HR}}(\mathtt{R}_0, \mathtt{C}_k) = \mathtt{R}_0^{-1}\mathtt{R}_k \tag{16}$$

The error functions are defined as

$$\begin{aligned}\mathbf{e}^{\mathrm{VO}}(\mathtt{C}_k, \mathtt{C}_{k+1}) &= \mathbf{z}_{k,k+1} \ominus \mathbf{h}^{\mathrm{VO}}(\mathtt{C}_k, \mathtt{C}_{k+1}) \\ &= \mathtt{t2v}\left(\{\mathbf{h}^{\mathrm{VO}}(\mathtt{C}_k, \mathtt{C}_{k+1})\}^{-1}\mathtt{T}_{k,k+1}\right) \quad (17)\\ \mathbf{e}^{\mathrm{HR}}(\mathtt{R}_0, \mathtt{C}_k) &= \mathbf{z}_{0k} \ominus \mathbf{h}^{\mathrm{HR}}(\mathtt{R}_0, \mathtt{C}_k) \\ &= \mathtt{t2v}\left(\{\mathbf{h}^{\mathrm{HR}}(\mathtt{R}_0, \mathtt{C}_k)\}^{-1}\mathtt{R}_{0k}\right) \quad (18)\end{aligned}$$

where function $\mathtt{t2v} : \mathrm{SO}(3) \rightarrow \mathbb{R}^3; \mathrm{SE}(3) \rightarrow \mathbb{R}^6$ converts rotation matrix to unit quaternion vector for a particular rotation, while keeps the translation vector remaining unchanged. For loosely-coupled VO, measurement information matrix $\mathbf{\Omega}^{\mathrm{VO}}$ between neighboring frames is modeled as constant in this work.

B. Tightly-Coupled Model

The tightly-coupled model is demonstrated in Fig. 2(b). Besides heading measurements in the loosely-coupled model, another two types of measurements have been considered: Landmark projection $\mathbf{z}_{\mathrm{L}ik}$ measures 3D-2D projection for landmark i at time k, while landmark location $\mathbf{z}_{\mathrm{L}i\mathrm{L}j}$ constrains three-dimensional distance between landmarks $\mathbf{x}_{\mathrm{L}i}$ and $\mathbf{x}_{\mathrm{L}j}$.

Landmark i in 3D are parameterized as $\mathbf{x}_{\mathrm{L}i} = [x_{\mathrm{L}i}, y_{\mathrm{L}i}, z_{\mathrm{L}i}]^\top$. According to epipolar geometry, projected coordinates v on image planes are identical for parallel cameras. Thus we extend the measurement equation (2) to stereo case such that

$$\mathbf{h}^{\mathrm{LM}}(\mathbf{x}_{\mathrm{L}i}, \mathtt{C}_k) = \begin{bmatrix} u_{i1} & v_{i1} & u_{i2} \end{bmatrix}^\top \tag{19}$$

where

$$\begin{bmatrix} u_{i1} \\ v_{i1} \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & c_u \\ 0 & f & c_v \\ 0 & 0 & 1 \end{bmatrix} \left(\mathtt{C}_k \begin{bmatrix} x_{\mathrm{L}i} \\ y_{\mathrm{L}i} \\ z_{\mathrm{L}i} \\ 1 \end{bmatrix} \right) \tag{20}$$

$$\begin{bmatrix} u_{i2} \\ v_{i2} \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & c_u \\ 0 & f & c_v \\ 0 & 0 & 1 \end{bmatrix} \left(\mathtt{C}_k \begin{bmatrix} x_{\mathrm{L}i} \\ y_{\mathrm{L}i} \\ z_{\mathrm{L}i} \\ 1 \end{bmatrix} - \begin{bmatrix} s \\ 0 \\ 0 \end{bmatrix} \right) \tag{21}$$

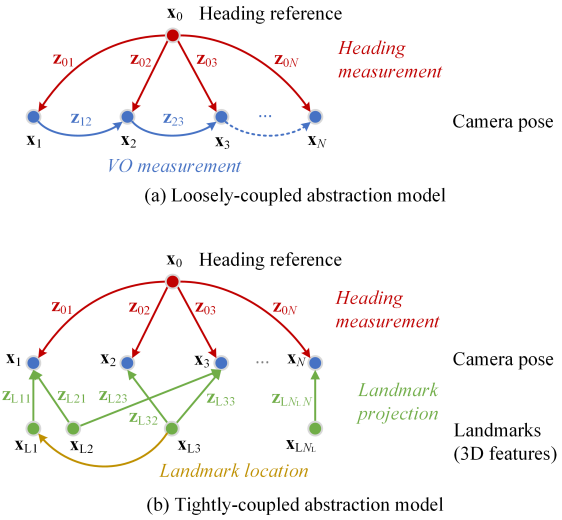


Fig. 2. Abstraction of measurement and graph formulation.

and $\mathtt{C}_k = \mathtt{v2t}(\mathbf{x}_k) = \left[\mathtt{R}_k | \mathbf{t}_k\right]$. Parameters f, c_u, c_v, s are camera-dependent and have been explained in Section III-A. The error function for landmark projection is defined as

$$\mathbf{e}^{\mathrm{LM}}(\mathbf{x}_{\mathrm{L}i}, \mathtt{C}_k) = \mathbf{z}_{\mathrm{L}ik} - \mathbf{h}^{\mathrm{LM}}(\mathbf{x}_{\mathrm{L}i}, \mathtt{C}_k) \tag{22}$$

where $\mathbf{z}_{\mathrm{L}ik} = [\tilde{u}_{i1}, \tilde{v}_{i1}, \tilde{u}_{i2}]^\top$ denotes projected coordinates vector for the stereo pair.

Sometimes landmark location constraints are desired to represent spatial relation among landmarks. The error function for landmark location is

$$\mathbf{e}^{\mathrm{LC}}(\mathbf{x}_{\mathrm{L}i}, \mathbf{x}_{\mathrm{L}j}) = \mathbf{z}_{\mathrm{L}i\mathrm{L}j} - \mathbf{h}^{\mathrm{LC}}(\mathbf{x}_{\mathrm{L}i}, \mathbf{x}_{\mathrm{L}j}) \tag{23}$$

where $\mathbf{z}_{\mathrm{L}i\mathrm{L}j} \in \mathbb{R}^3$ and $\mathbf{h}^{\mathrm{LC}}(\mathbf{x}_{\mathrm{L}i}, \mathbf{x}_{\mathrm{L}j}) = \mathbf{x}_{\mathrm{L}i} - \mathbf{x}_{\mathrm{L}j}$. In tightly-coupled model, information matrices $\mathbf{\Omega}^{\mathrm{LM}}$ and $\mathbf{\Omega}^{\mathrm{LC}}$ are both set to diagonal matrices.

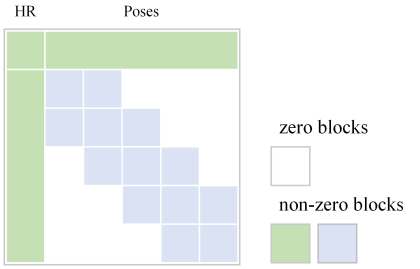
C. Structure of the Abstraction Model

Optimization-based state estimation problem had become computationally feasible since the sparse structure of the problem was discovered. The proposed abstraction model brings new types of constraints such as heading measurements, which may cause weaker sparsity of the abstraction model. As the heading reference vertex can be deemed as a subset of camera pose vertices, Jacobian in (8) has the form as

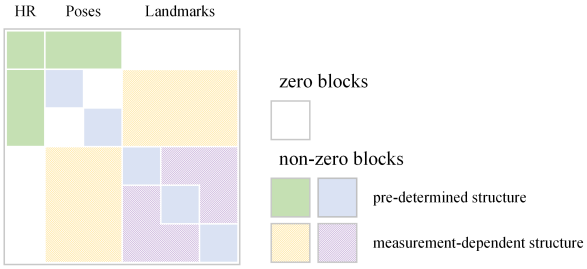
$$\mathbf{J}_{ij} = \begin{bmatrix} \cdots & \underbrace{\dfrac{\partial \mathbf{e}_{ij}}{\partial \mathbf{x}_i}}_{i\text{th element}} & \cdots & \underbrace{\dfrac{\partial \mathbf{e}_{ij}}{\partial \mathbf{x}_j}}_{j\text{th element}} & \cdots \end{bmatrix} \tag{24}$$

where all zero items are omitted. The structure of matrix $\mathbf{H}_{ij}$ is

$$\mathbf{H}_{ij} = \begin{bmatrix} \ddots & & & & \\ & \frac{\partial \mathbf{e}_{ij}}{\partial \mathbf{x}_i}^\top \mathbf{\Omega}_{ij} \frac{\partial \mathbf{e}_{ij}}{\partial \mathbf{x}_i} & \cdots & \frac{\partial \mathbf{e}_{ij}}{\partial \mathbf{x}_i}^\top \mathbf{\Omega}_{ij} \frac{\partial \mathbf{e}_{ij}}{\partial \mathbf{x}_j} & \\ & \vdots & & \vdots & \\ & \frac{\partial \mathbf{e}_{ij}}{\partial \mathbf{x}_j}^\top \mathbf{\Omega}_{ij} \frac{\partial \mathbf{e}_{ij}}{\partial \mathbf{x}_i} & \cdots & \frac{\partial \mathbf{e}_{ij}}{\partial \mathbf{x}_j}^\top \mathbf{\Omega}_{ij} \frac{\partial \mathbf{e}_{ij}}{\partial \mathbf{x}_j} & \\ & & & & \ddots \end{bmatrix} \tag{25}$$



(a) Structure of matrix **H** in loosely-coupled abstraction model



(b) Structure of matrix **H** in tightly-coupled abstraction model

Fig. 3. Illustration of $\mathbf{H}$ matrices for the loosely-coupled and tightly-coupled model, where "heading reference" is abbreviated as "HR". In (a), block tridiagonal matrix is obtained in lower right part due to constraints between consecutive VO poses; In (b), non-zero blocks exist mainly due to landmark projection measurements. Besides, landmark location constraints may cause non-zero blocks in lower right part but the sparsity will remain as the number of landmark location constraints is small compared to the total number of landmarks.

and matrix $\mathbf{H}$ has the structure shown in Fig. 3, from which we know that the proposed graph model remains sparse. The newly-introduced heading measurements and VO measurements lead to block tridiagonal matrices in the loosely-coupled model. In the tightly-coupled model, landmark location constraints may slightly weaken the sparsity of generated graph, but we are not able to predict the sparsity unless measurements have been taken. Selection of abstract level is problem-dependent by considering sensors, measurements, constraints and the computing power. Comparative analysis between these two models can be found in Section VI and VII.

D. Vertex Removal in Abstraction Model

In above-proposed model, hundreds of landmark vertices are generated each frame in the tightly-coupled model. Even in the loosely-coupled model, vertex number increases at linear rate. It is impracticable to retain all vertices and edges during real-time optimization process. As the optimization problem is nonlinear and non-convex, only qualitative discussion with theoretical assumptions will be presented here to help better explain simulative and experimental results.

Given a graph model with a local minimum $\mathbf{x}^* = \arg\min_{\mathbf{x}} \sum_{<i,j>\in\mathcal{C}} {\mathbf{e}_{ij}}^\top \mathbf{\Omega}_{ij} \mathbf{e}_{ij}$, we consider an augmented optimization problem

$$\mathbf{x}^*_{\text{aug}} = \arg\min_{\mathbf{x}_{\text{aug}}} \left(\sum_{<i,j>\in\mathcal{C}} {\mathbf{e}_{ij}}^\top \mathbf{\Omega}_{ij} \mathbf{e}_{ij} + \sum_{<i,j>\in\mathcal{C}_{\text{new}}} {\mathbf{e}_{ij}}^\top \mathbf{\Omega}_{ij} \mathbf{e}_{ij} \right) \quad (26)$$

where $\mathbf{x}_{\text{aug}} = \left[\mathbf{x}^\top, \mathbf{x}_{\text{new}}^\top\right]^\top$, and $\mathcal{C}_{\text{new}}$ only contains indices connecting to $\mathbf{x}_{\text{new}}$.

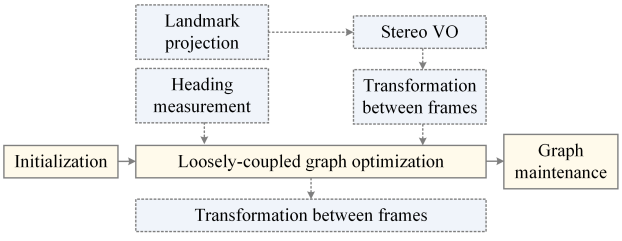
Let us expand the error terms as $\mathbf{e}_{ij} = \mathbf{e}_{ij}(\mathbf{z}_{ij}, \mathbf{x}_i, \mathbf{x}_j) = \mathbf{e}_{ij}(\mathbf{x}_i, \mathbf{x}_j)$, the cost function in augmented problem can be seperated to two parts, where each is a quadratic form w.r.t. $\mathbf{e}_{ij}(\mathbf{x}_i, \mathbf{x}_j)$. It is obvious that for the augmented problem, $\mathbf{x}^*_{\text{aug}} = \left[\mathbf{x}^{*\top}, \mathbf{x}^{*\top}_{\text{new}}\right]^\top$ is a local minimum where

$$\mathbf{x}^*_{\text{new}} = \arg\min_{\mathbf{x}_{\text{new}}} \sum_{<i,j>\in\mathcal{C}_{\text{new}}} {\mathbf{e}_{ij}}^\top \mathbf{\Omega}_{ij} \mathbf{e}_{ij} \quad (27)$$

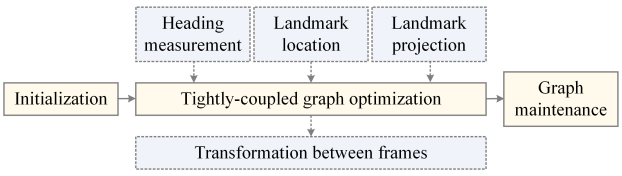
Whether the local minimum is also the global minimum depends on the error function $\mathbf{e}_{ij}$. In a trivial case where $\mathbf{e}_{ij}$ is linear w.r.t. $\mathbf{x}_i$ and $\mathbf{x}_j$, the augmented problem is convex, which makes optimization effortless. Unfortunately, as discussed in Section IV-A and IV-B, most error functions in pose estimation are nonlinear and non-convex. Thus the initial value in optimization should be closely selected such that the numerical method is less like to fall into the local optimum.

V. Heading Reference Assisted Pose Estimation (HRPE)

As illustrated in Fig. 4, this section presents the pose estimation scheme, where landmarks on stereo images and headings are regarded as raw measurements from cameras and heading reference sensor, respectively. For loosely-coupled model, egomotion estimation based on image feature correspondence (SVO) has to be performed first to obtain VO measurements. For tightly-coupled model, the matched feature points are directly abstracted as vertices in pose graph. At each time when measurements have been acquired, the updated graph is optimized w.r.t. the cost function, then vertices and edges out of the sliding window are discarded during graph maintenance. For convenient representation, we label pose estimation approaches based on the loosely-coupled and tightly-coupled model as "LC-HRPE" and "TC-HRPE" in subsequent sections.



(a) Structure of LC-HRPE



(b) Structure of TC-HRPE

Fig. 4. Heading reference assisted pose estimation scheme. Solid arrows represent evolution over time for a single estimation, while dotted arrows denote information flow.

A. Initialization

Initialization aims to acquire an accurate estimation of the heading reference. By taking advantage of insignificant VO drift, the heading reference and camera poses in the first few frames are optimized simultaneously to minimize the influence of heading measurement random errors. During the rest of the optimization process, the heading reference vertex is set to fixed.

B. Graph Optimization

As the frequency of heading measurements is usually much higher than cameras, filtering and downsampling can be done in preprocessing for noise reduction and data synchronization. It is assumed that all measurements are synchronized before graph optimization. The optimization goals in LC-HRPE and TC-HRPE are written as

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \sum_{<i,j>\in\mathcal{C}^{\mathrm{VO}}} \mathbf{e}^{\mathrm{VO}\top}\boldsymbol{\Omega}^{\mathrm{VO}}\mathbf{e}^{\mathrm{VO}} + \sum_{<i,j>\in\mathcal{C}^{\mathrm{HR}}} \mathbf{e}^{\mathrm{HR}\top}\boldsymbol{\Omega}^{\mathrm{HR}}\mathbf{e}^{\mathrm{HR}} \quad (28)$$

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \sum_{<i,j>\in\mathcal{C}^{\mathrm{HR}}} \mathbf{e}^{\mathrm{HR}\top}\boldsymbol{\Omega}^{\mathrm{HR}}\mathbf{e}^{\mathrm{HR}} + \sum_{<i,j>\in\mathcal{C}^{\mathrm{LM}}} \mathbf{e}^{\mathrm{LM}\top}\boldsymbol{\Omega}^{\mathrm{LM}}\mathbf{e}^{\mathrm{LM}} + \sum_{<i,j>\in\mathcal{C}^{\mathrm{LC}}} \mathbf{e}^{\mathrm{LC}\top}\boldsymbol{\Omega}^{\mathrm{LC}}\mathbf{e}^{\mathrm{LC}} \quad (29)$$

respectively, where $\mathcal{C}^{\mathrm{VO}}$, $\mathcal{C}^{\mathrm{HR}}$, $\mathcal{C}^{\mathrm{LM}}$ and $\mathcal{C}^{\mathrm{LC}}$ denote index sets containing measurements of VO, heading, landmark projection and landmark location; Subscripts in error functions are omitted for convenient representation. In this work, each landmark vertex will be projected to a single camera pose vertex. Landmark locations are regarded identical if a landmark is being tracked constantly, and we may set $\boldsymbol{\Omega}^{\mathrm{LC}}$ large to represent identical vertices.

By considering the non-convexity of the optimization problem, the initial pose estimates are not set to identity matrices but VO estimated poses such that local minimum could be avoided. In TC-HRPE, since we are not estimating structure from motion, all feature vertices are fixed to avoid unnecessary optimization on 3D feature points.

C. Maintenance of the Dynamic Graph

The real-time performance of state estimation depends extensively on the graph size. In SLAM problem, post-processing is acceptable to exploit all measurements, while it is not the case for VO. Although some approaches have been proposed for reducing computational load by minimizing graph size [30], [31], we implement the straightforward sliding window to restrain unlimited growth of computational load. During graph maintenance, all vertices and their edges out of the sliding window will be dropped out.

We have discussed the influence of vertex removal theoretically in Section IV-D. If all current vertices are optimal, the graph optimization problem can be simplified by deleting all edges that are not adjacent to current vertex. However, a sliding window is essential in our proposed model because i) vertices are never optimal with the existence of measurement noise; ii) landmark location measurements in TC-HRPE require simultaneous existence of several pose vertices.

VI. Simulative Studies

The KITTI odometry dataset has been selected to run simulative tests. Since no heading reference is available in the dataset, orientation measurements from OXTS RT3003 Inertial and GPS Navigation System with manually added Gaussian noises $\mathcal{N}\left(\mathbf{0}, \mathrm{diag}(\sigma_r^2, \sigma_p^2, \sigma_y^2)\right)$ for roll, pitch, and yaw are regarded as heading measurements. In all simulations and experiments, *libviso2* [24] and *g2o* [28] are used for VO implementation and graph optimization, respectively.

Simulation results with sliding window size 10 and Euler angle noise parameters $\sigma_r = \sigma_p = \sigma_y = 5$ degrees are listed in Table I, where only several sequences with relative long traveling distances are selected during evaluation to demonstrate pose estimation results of the proposed approach with large rotational drifts. As this work considers pose estimation for ground vehicles, translation error on camera frame $x - z$ plane and rotation error on yaw are focused. Results show that the proposed approach improves VO by reducing translational and rotational pose estimation errors and their standard deviations. In particular, yaw estimation errors for all evaluated sequences are shown in Fig. 5. It is hard to fully compensate drift error in VIO since IMU also suffers from bias error inherently [9]. Compared with VIO which tend to be unbounded in estimation error [32], [33], it is evident that the rotation errors are bounded with the assistance of heading measurement. Moreover, we notice that TC-HRPE has more robust estimation performance on rotation on account of its smaller error standard deviations.

The noise level of heading measurements and sliding window size do play an essential role in algorithm performance. Next, we discuss accuracy and efficiency w.r.t. sliding window size and heading measurement error.

A. Accuracy w.r.t. Heading Measurement Error

For a particular heading reference sensor, since all measurements are represented in sensor local frame, the system error caused by disalignment between sensor and vehicle body frame has no effect on estimation accuracy. To investigate the influence of random error, Fig. 6 shows the changing trend while the standard deviations of roll, pitch, yaw measurement errors vary from 5 to 20 degrees. It is observed that accurate heading measurements generally lead to superior rotation estimation. The proposed estimation approach reduces rotation error compared to raw heading measurements with noises. The standard deviation of estimation error increases with inferior heading measurements in most cases.

B. Accuracy w.r.t. Sliding Window Size

As shown in Fig. 7, there is no significant difference observed with varying sliding window size in simulation. As discussed in Section IV-D, dropping out previous vertices and their adjacent edges will not influence current estimation in ideal case. However, to make sure the heading reference measurements takes effects in the abstraction model and stabilize the graph in practice, size 10 is appropriate in this work.

For tightly-coupled model, increasing sliding window size enables the possibility to consider more features that are

Table I. Simulation results of SVO, Loosely-Coupled HRPE (LC-HRPE) and Tightly-Coupled HRPE (TC-HRPE).

Sequence	Travelling Distance	SVO: Trans(m)/Yaw(deg)		LC-HRPE: Trans(m)/Yaw(deg)		TC-HRPE: Trans(m)/Yaw(deg)	
		Avg Error	Std Dev	Avg Error	Std Dev	Avg Error	Std Dev
KITTI 00	3,724 m	34.56/10.21	27.76/5.60	10.39/1.49	7.65/1.12	20.79/0.91	10.73/0.87
KITTI 02	5,067 m	61.00/10.38	45.11/5.37	4.10/1.46	1.72/1.14	9.15/0.95	3.21/0.66
KITTI 05	2,205 m	22.32/5.93	17.46/3.32	19.61/1.61	10.34/1.25	15.14/1.98	7.09/1.45
KITTI 08	3,222 m	58.29/7.45	39.11/5.01	23.75/1.64	12.81/1.26	27.62/0.58	11.65/0.42

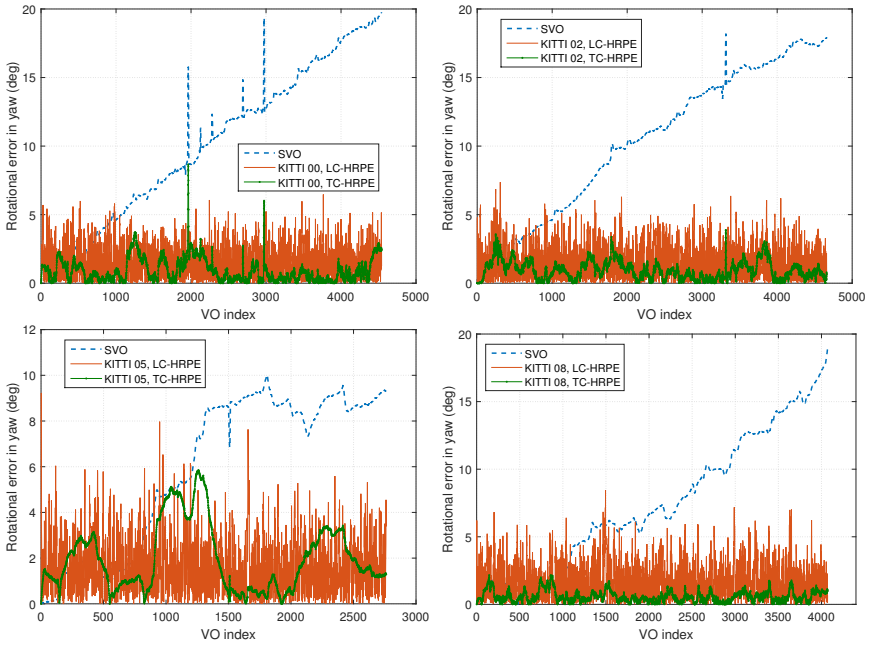


Fig. 5. Yaw estimation error of sequence KITTI 00, 02, 05 and 08 w.r.t. VO index, with heading measurement error $\sigma_r = \sigma_p = \sigma_y = 5$ degrees and sliding window size 10.

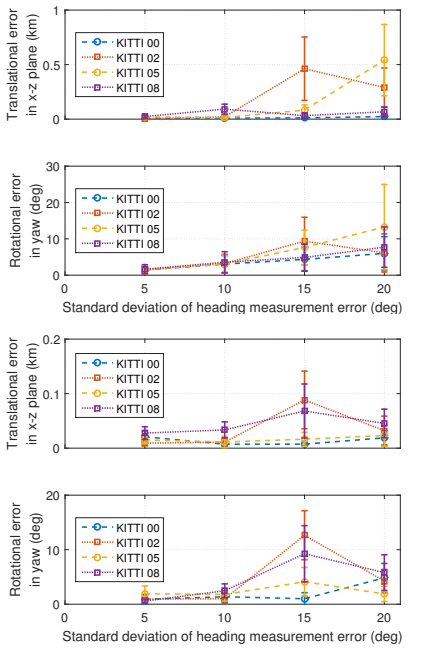


Fig. 6. Estimation accuracy w.r.t. heading measurement error using loosely-coupled (upper figure) and tightly-coupled model (lower figure), respectively.

observable in multiple frames. Furthermore, a relatively long sliding window may help retain more information such that it is possible to achieve global optimization and mapping when real-time estimation is not required.

C. Time Consumption w.r.t. Sliding Window Size

Sliding window size will definitely influence the efficiency while optimizing the graph. In this work, we explore the time consumed, and results are demonstrated in Fig. 8. Maximum of two features are allowed in each bucket with width and height 50×50 in VO. For evaluated four KITTI sequences, we compute the executing time by averaging time per pose, given different heading measurement noise with standard deviations 5, 10, 15, and 20 degrees. The average executing time per pose for LC-HRPE and TC-HRPE are 0.06 second and 0.125 second with window size 10. Since features are detected and extracted for each new image frame in real time, time for feature detection and extraction can be deemed as constant complexity, thus time for optimization occupies a larger proportion of overall time consumption as sliding window size becomes larger. By selecting appropriate abstraction model and sliding window size according to hardware performance, real-time implementation is attainable.

VII. Experimental Results

To test the proposed approach further, we have conducted experimental studies using our self-collected sequences near Nanyang Technological University (NTU) campus. The sequences contain challenging driving scenarios including passing humps, image intensity variation and frequent acceleration/deceleration (see Fig. 9 for graphical illustrations).

A. Experimental Platform

The experimental platform for data collection is shown in Fig. 10. The stereo vision system that consists of two Fla3 image sensors (FL3-U3-13Y3M-C) and Kowa 1/2 inch lenses has been built to collect stereo images at the resolution of 640×422 with baseline 0.3 meter. Heading measurements w.r.t. NED frame are obtained from DJI A3 GPS-IMU module, which measures earth magnetic field and fuses it with internally-measured linear acceleration and angular velocity. The stereo images are recorded at 40 frames per second but have been down-sampled to 10 Hz. Fused translation poses from DJI A3 GPS-IMU module are used as vehicle's groundtruth position at 50 Hz. To evaluate the proposed approach, the groundtruth is synchronized with estimated pose using the ROS `Synchronizer` filter templated on `approximateTime` policy. The information matrices for VO (LC-HRPE), heading measurements (LC-HRPE), same

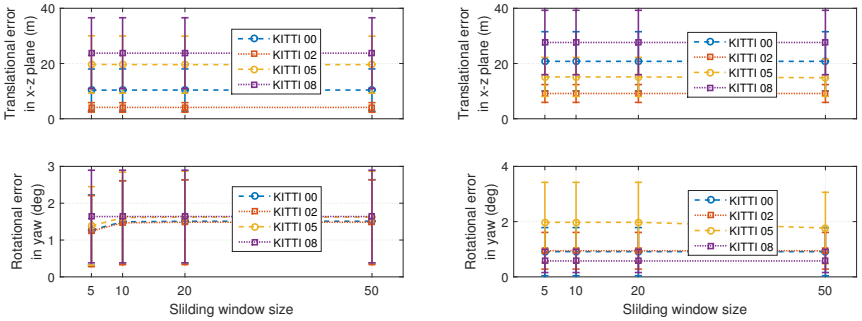


Fig. 7. Estimation accuracy w.r.t. sliding window size using loosely-coupled (left) and tightly-coupled model (right) with heading measurement error $\sigma_r = \sigma_p = \sigma_y = 5$ degrees.

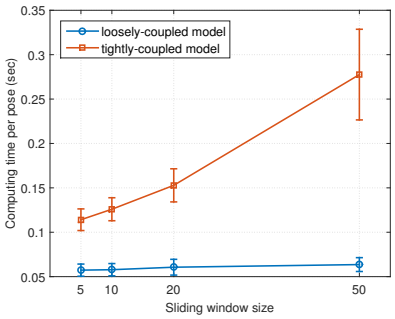


Fig. 8. Average computing time per pose for simulated sequences.

Fig. 9. Outliers at turning (left), passing humps (middle), and intensity variation (right) in NTU sequences 01-04.

Fig. 10. Our experimental platform.

feature constraint (TC-HRPE), feature-pose constraint (TC-HRPE) and heading measurements (TC-HRPE) are $2000\mathbf{I}_6$, $100\mathbf{I}_3$, $10000\mathbf{I}_3$ $0.001\mathbf{I}_2$ and $2000\mathbf{I}_3$, respectively.

B. Pose Estimation Performance

We only evaluate translation error quantitatively because orientations have been used as heading measurements thus cannot be taken as fair groundtruth. The qualitative performance of rotation estimation is still accessible from vehicle trajectories. The pose estimation performance for LC-HRPE and TC-HRPE is quantitatively shown in Table. II, where results demonstrate that the translation estimation error has been substantially reduced with the assistance of heading measurements. Specifically, the average translation errors for SVO, LC-HRPE and TC-HRPE are 153.85 meters, 24.29 meters and 23.80 meters. The robustness of estimation has been improved by the proposed approach according to evidently smaller error standard deviation.

Fig. 11 and Fig. 12 illustrate comparative results on vehicle's trajectories and translation estimation error for all four sequences. Rotation error increases rapidly, especially when the vehicle is making turns in environments with outliers, illumination variation, and poor features. For the same reason, translation error tends to be unbounded with SVO only. The proposed approach slows translation error growth rate such that the position is still effective in case of large rotation error in VO. Although it is extremely difficult to bound translation error due to egomotion iteration and error accumulation, accurate rotation estimation contributes significantly to VO performance.

C. Real-time Performance

The proposed system has been evaluated on a desktop computer with an Intel Core i7-4770 CPU@3.40GHz and 12GB memory. Experimental real-time performance are listed in Table II, which shows that the average execution frequencies for SVO, LC-HRPE and TC-HRPE are 20.8 Hz, 19.7 Hz and 11.1 Hz, respectively. The loosely-coupled model brings a little extra computational load to pure odometry, compared to the tightly-coupled model that doubles operation time. For TC-HRPE, not only sliding window size but the feature detection and extraction parameters will largely influence real-time performance. In our experiments, the average graph edge number at each iteration for TC-HRPE can be found in Fig.13. Lowering feature detection and selection threshold may help increase VO performance but will lead to delayed results.

VIII. Conclusion and Future Work

In this work, we have presented a stereo pose estimation framework that needs assistance from heading reference, with elaboration on the proposed abstraction model and evaluations in both public and self-collected datasets. The proposed

Table II. Experimental results of SVO, Loosely-Coupled HRPE (LC-HRPE) and Tightly-Coupled HRPE (TC-HRPE).

Sequence	Travelling Distance	SVO: Trans(m)			LC-HRPE: Trans(m)			TC-HRPE: Trans(m)		
		Avg Error	Std Dev	Frequency	Avg Error	Std Dev	Frequency	Avg Error	Std Dev	Frequency
NTU 01	1,244 m	132.32	115.48	22.6 Hz	22.3	10.62	20.8 Hz	14.85	10.85	11.5 Hz
NTU 02	1,249 m	184.57	120.00	20.3 Hz	19.53	8.13	19.0 Hz	35.62	9.41	10.8 Hz
NTU 03	860 m	163.44	122.68	20.3 Hz	21.41	8.19	19.7 Hz	28.20	12.24	10.9 Hz
NTU 04	1,204 m	141.98	107.64	19.9 Hz	31.96	26.76	19.4 Hz	19.30	14.62	11.2 Hz
Total/Avg	4,557 m	153.85	-	20.8 Hz	24.29	-	19.7 Hz	23.80	-	11.1 Hz

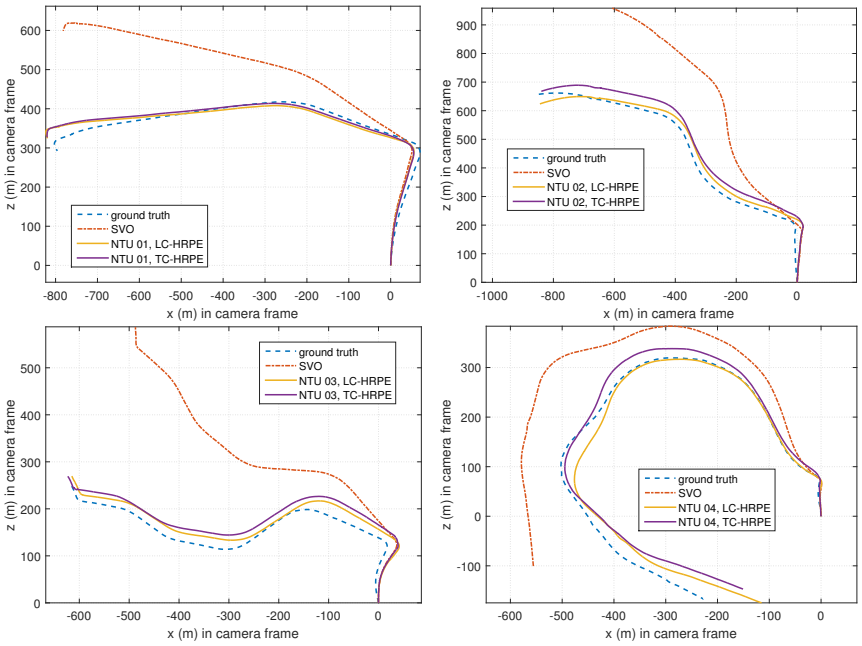


Fig. 11. Experimental results based on self-collected sequences NTU 01-04.

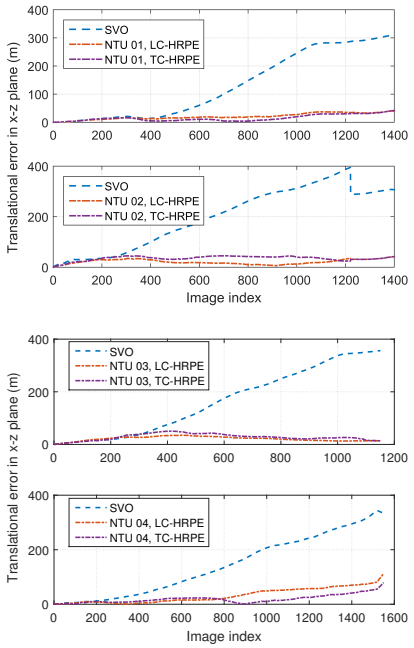


Fig. 12. Translation estimation error based on NTU 01-04.

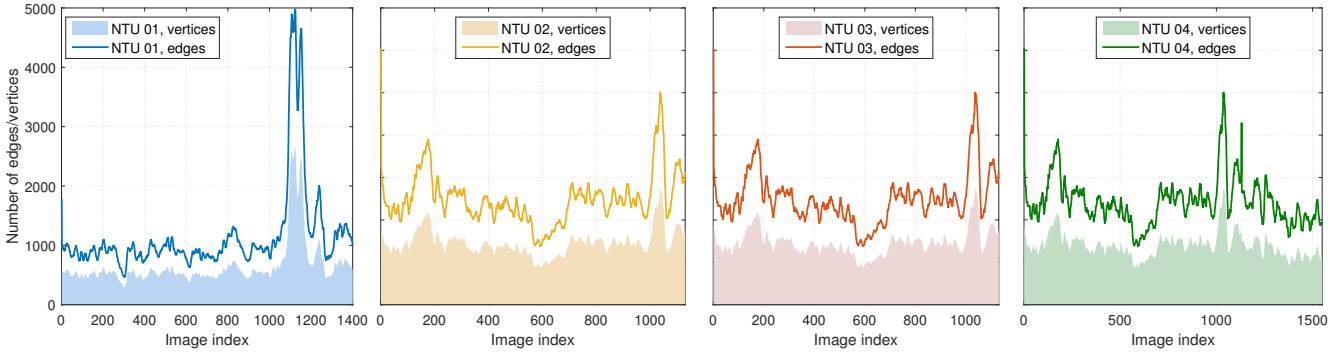


Fig. 13. The sizes of graphs to be optimized based on tightly-coupled model and self-collected sequences NTU 01-04.

system has demonstrated its effectiveness and robustness by significantly reducing translation and rotation estimation error, compared to pure stereo visual odometry. The extra computational cost is acceptable for real-time requirement of pose estimation. The absolute heading reference bounds the rotation drift, whereas the translation drift will increase unboundedly at a significantly lower rate. As all other visual pose estimation approaches, poor illumination conditions and outliers in image features will cause performance degradation.

Sensor fusion is essential for autonomous vehicles to obtain robust pose estimation in varying scenarios. The proposed framework could be extended by catering to other types of sensors or combining measurements from multiple heading reference sensors. In the front-end, various image features and outliers rejection techniques can be selected and implemented to make the system adaptive to complex egomotion and illumination conditions. Future exploration may also include incorporating the tracking into mapping and loop closing such that a heading reference assisted SLAM system would be built.

ACKNOWLEDGMENT

This work is supported by Singapore Technologies Kinetics Ltd. The authors would like to thank the editors and anonymous reviewers for their valuable comments to improve the quality of this paper.

REFERENCES

[1] Y. Cheng, M. Maimone, and L. Matthies, "Visual odometry on the mars exploration rovers," in *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, vol. 1, pp. 903–910, IEEE, 2005.

[2] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry for ground vehicle applications," *Journal of Field Robotics*, vol. 23, no. 1, pp. 3–20, 2006.

[3] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[4] S. Shen, Y. Mulgaonkar, N. Michael, and V. Kumar, "Multi-sensor fusion for robust autonomous flight in indoor and outdoor environments with a rotorcraft MAV," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pp. 4974–4981, IEEE, 2014.

[5] J. Zhang and S. Singh, "Visual-lidar odometry and mapping: Low-drift, robust, and fast," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pp. 2174–2181, IEEE, 2015.

[6] M. Agrawal and K. Konolige, "Real-time localization in outdoor environments using stereo vision and inexpensive gps," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 3, pp. 1063–1068, IEEE, 2006.

[7] R. Jiang, S. Yang, S. S. Ge, H. Wang, and T. H. Lee, "Geometric map-assisted localization for mobile robots based on uniform-gaussian distribution," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 789–795, 2017.

[8] I. P. Alonso, D. F. Llorca, M. Gavilán, S. Á. Pardo, M. Á. García-Garrido, L. Vlacic, and M. Á. Sotelo, "Accurate global localization using visual odometry and digital maps on urban environments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1535–1545, 2012.

[9] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct ekf-based approach," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pp. 298–304, IEEE, 2015.

[10] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint kalman filter for vision-aided inertial navigation," in *Robotics and automation, 2007 IEEE international conference on*, pp. 3565–3572, IEEE, 2007.

[11] J. Lobo and J. Dias, "Vision and inertial sensor cooperation using gravity as a vertical reference," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1597–1608, 2003.

[12] K. Wang, Y.-H. Liu, and L. Li, "A simple and parallel algorithm for real-time robot localization by fusing monocular vision and odometry/ahrs sensors," *IEEE/ASME Transactions on Mechatronics*, vol. 19, no. 4, pp. 1447–1457, 2014.

[13] J. M. Falquez, M. Kasper, and G. Sibley, "Inertial aided dense & semi-dense methods for robust direct visual odometry," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, pp. 3601–3607, IEEE, 2016.

[14] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual–inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.

[15] T. Lupton and S. Sukkarieh, "Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions," *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 61–76, 2012.

[16] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual–inertial odometry," *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, 2017.

[17] P. Piniés, T. Lupton, S. Sukkarieh, and J. D. Tardós, "Inertial aiding of inverse depth slam using a monocular camera," in *Robotics and Automation, 2007 IEEE International Conference on*, pp. 2797–2802, IEEE, 2007.

[18] M. Li and A. I. Mourikis, "High-precision, consistent ekf-based visual-inertial odometry," *The International Journal of Robotics Research*, vol. 32, no. 6, pp. 690–711, 2013.

[19] F. Santoso, M. A. Garratt, and S. G. Anavatti, "Visual–inertial navigation systems for aerial robotics: Sensor fusion and technology," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 1, pp. 260–275, 2017.

[20] H. Liu, F. Sun, B. Fang, and X. Zhang, "Robotic room-level localization using multiple sets of sonar measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 1, pp. 2–13, 2017.

[21] Z. Yang and S. Shen, "Monocular visual–inertial state estimation with online initialization and camera–imu extrinsic calibration," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 1, pp. 39–51, 2017.

[22] O. Saurer, P. Vasseur, R. Boutteau, C. Demonceaux, M. Pollefeys, and F. Fraundorfer, "Homography based egomotion estimation with a common direction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 2, pp. 327–341, 2017.

[23] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *IEEE robotics & automation magazine*, vol. 18, no. 4, pp. 80–92, 2011.

[24] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pp. 963–968, IEEE, 2011.

[25] G. Wahba, "A least squares estimate of satellite attitude," *SIAM review*, vol. 7, no. 3, pp. 409–409, 1965.

[26] M. D. Shuster and S. D. Oh, "Three-axis attitude determination from vector observations," *Journal of Guidance, Control, and Dynamics*, vol. 4, no. 1, pp. 70–77, 1981.

[27] M. Kaess, A. Ranganathan, and F. Dellaert, "isam: Fast incremental smoothing and mapping with efficient data association," in *Robotics and Automation, 2007 IEEE International Conference on*, pp. 1670–1677, IEEE, 2007.

[28] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g^2o: A general framework for graph optimization," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pp. 3607–3613, IEEE, 2011.

[29] K. Konolige and W. Garage, "Sparse sparse bundle adjustment.," in *BMVC*, vol. 10, pp. 102–1, Citeseer, 2010.

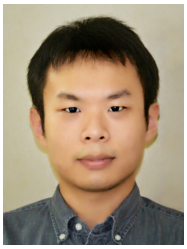
[30] N. Carlevaris-Bianco, M. Kaess, and R. M. Eustice, "Generic node removal for factor-graph slam," *IEEE Transactions on Robotics*, vol. 30, no. 6, pp. 1371–1385, 2014.

[31] N. Carlevaris-Bianco and R. M. Eustice, "Generic factor-based node marginalization and edge sparsification for pose-graph slam," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pp. 5748–5755, IEEE, 2013.

[32] V. Usenko, J. Engel, J. Stückler, and D. Cremers, "Direct visual-inertial odometry with stereo cameras," in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*, pp. 1885–1892, IEEE, 2016.

[33] Y. Liu, R. Xiong, Y. Wang, H. Huang, X. Xie, X. Liu, and G. Zhang, "Stereo visual-inertial odometry with multiple kalman filters ensemble," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 10, pp. 6205–6216, 2016.

Han Wang is currently an Associate Professor with the School of Electrical and Electronics Engineering, Nanyang Technological University. He received his Bachelor degree in Computer Science from Northeast Heavy Machinery Institute (China), and Ph.D. degree from the University of Leeds (UK), respectively. His research interests include computer vision and robotics. He has published over 120 top quality international conference and journal papers. He has been invited as a member of Editorial Advisory Board, the Open Electrical and Electronic Engineering Journal. Dr. Wang is a senior member of IEEE.



Rui Jiang received the B.Eng. degree in Measurement, Control Technique and Instruments from Harbin Institute of Technology, Harbin, China, in 2014. He is currently a project officer with the School of Electrical and Electronics Engineering, Nanyang Technological University and pursuing his Ph.D. degree with the department of Electrical and Computer Engineering, National University of Singapore. His research interests include intelligent sensing, localization and navigation for autonomous vehicles.

Handuo Zhang is currently a Ph.D. student in the School of Electrical and Electronics Engineering, Nanyang Technological University since 2016. He received his Bachelor degree in Automation and Master degree in Pattern Recognition and Intelligent System both from Northeastern University, China. He was an assistant researcher in the Shenyang Institute of Automation Chinese Academy of Sciences (2013-2015). His research interests include 3D reconstruction in large-scale environment, robot perception and visual SLAM.

Shuzhi Sam Ge received the B.Sc. degree from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1986, and the Ph.D. degree from the Imperial College of Science, Technology and Medicine, University of London, London, U.K., in 1993. He is the Founding Director of the Social Robotics Laboratory with the Interactive Digital Media Institute, National University of Singapore, Singapore, where he is a Professor with the Department of Electrical and Computer Engineering. Prof. Ge is the Editor-in-Chief of the International Journal of Social Robotics. He has served as an Associate Editor for a number of flagship journals. He serves as an Editor of the Automation and Control Engineering Series (Taylor & Francis). He also served as the Vice President of the technical activities from 2009 to 2010, membership activities from 2011 to 2012, and the IEEE Control Systems Society.