IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 23, 2020, accepted June 27, 2020, date of publication July 1, 2020, date of current version July 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3006210

Multi-Agent Collaborative GNSS/Camera/INS Integration Aided by Inter-Ranging for Vehicular Navigation in Urban Areas

WEISONG WEN[1], XIWEI BAI[1], GUOHAO ZHANG[1], SHENGDONG CHEN[2], FENG YUAN[2], AND LI-TA HSU[1], (Member, IEEE)

[1]Intelligent Positioning and Navigation Laboratory (IPNL), The Hong Kong Polytechnic University, Hong Kong, China
[2]Institute of Software Application Technology, Guangzhou Chinese Academy of Sciences, Guangzhou 511458, China

Corresponding author: Li-Ta Hsu (lt.hsu@polyu.edu.hk)

This work was supported by the Hong Kong Polytechnic University Feng's Seed Fund (Security Enhancement of Positioning Sensors on Connect Autonomous Vehicles) under Grant P0013910 (ZVP9).

ABSTRACT Achieving accurate and reliable positioning in dynamic urban scenarios using low-cost vehicular onboard sensors, such as the global navigation satellite systems (GNSS), camera, and inertial measurement unit (IMU), is still a challenging problem. Multi-Agent collaborative integration (MCI) opens a new window for achieving this goal, by sharing the sensor measurements between multiple agents to further improve the accuracy of respective positioning. One of the major difficulties in MCI is to effectively connect all the sensor measurements arising from multiple independent agents. The popular approach is to find the overlapping areas between agents using active sensors, such as cameras. However, the performance of overlapping area detection is significantly degraded in outdoor urban areas due to the challenges arising from numerous unexpected moving objects and unstable illumination conditions. To fill this gap, this paper proposes to leverage both the camera-based overlapping area detection and the inter-ranging measurements to boost the cross-connection between multi-agents and brings the MCI to outdoor urban scenarios using low-cost onboard sensors. Moreover, a novel MCI framework is proposed to integrate the sensor measurements from the low-cost GNSS receiver, camera, IMU, and inter-ranging using state-of-the-art factor graph optimization (FGO) to fully explore their complementary properties. The proposed MCI framework is validated using two challenging datasets collected in urban canyons of Hong Kong. We conclude that the proposed MCI framework can effectively improve the positioning accuracy of the respective agents in the evaluated datasets. We believe that the proposed MCI framework has the potential to be prevalently adopted by the connected intelligent transportation systems (ITS) applications to provide robust positioning using low-cost onboard sensors in urban scenarios.

INDEX TERMS Multi-agent collaborative positioning, GNSS, camera, INS, inter-ranging, factor graph optimization, urban canyons.

I. INTRODUCTION

Robust, cost-effective, and accurate positioning is significant for the extensive commercialization of the emerging intelligent transportation systems (ITS) [1] with navigation requirements, such as advanced collision warning systems [2], speed advisory systems [3] and lane reservation systems [4]. The low-cost global navigation satellite systems (GNSS) receiver, inertial measurement units (IMU) and camera are the most ubiquitous sensor setups for vehicular systems, due to their cost-effectiveness. Unfortunately, achieving lane-level positioning in urban canyons using these onboard sensors is still a challenging problem. Decent accuracy with an error of fewer than 5 meters can be obtained using low-cost GNSS receiver in open areas with decent sky visibility. Whereas, the performance of GNSS solutions can be significantly degraded in urban canyons, due to the severe multipath effects and non-line-of-sight (NLOS) receptions caused by the high-rising building reflections and blockage [5]. To mitigate the impacts of the errors caused by multipath and

The associate editor coordinating the review of this manuscript and approving it for publication was Lubin Chang.

NLOS receptions, numerous researches [6]–[9] are proposed to exclude [8], [10], correct [9] or re-model [11] the outlier GNSS raw measurements to further improve the GNSS solution in urban canyons. However, these methods rely on the availability of additional costly 3D LiDAR sensors or 3D building model information. The performance of IMU relies on its cost and the bias corrections [12]–[14] from additional sensors, such as the GNSS receiver. The recently extensively studied visual-inertial navigation system (VINS) [15] can provide accurate and relative positioning in a short period. However, according to our findings in [16], the performance of VINS can be significantly degraded in urban canyons due to the numerous unexpected dynamic objects [17] and unstable illumination conditions. Therefore, the integration of GNSS/IMU/camera [18] is studied to make use of their complementary properties to provide better performance. However, its performance is still challenged in urban canyons. In short, achieving accurate positioning from the low-cost onboard sensors (GNSS/IMU/camera) from a single agent in the urban canyon is still a challenging problem to be solved.

Instead of estimating the vehicular state simply based on onboard sensors from a single agent, the multi-agents collaborative integration (MCI) opens a new window for improving the multi-agents positioning [19], [20] accuracy. The MCI makes use of the sensor measurements from multiple agents to facilitate respective positioning accuracy. One of the major difficulties of MCI is to effectively connect all the independent agents using intervehicle measurements, such as the inter-ranging measurement [21] which describes the relative distance between two agents. Then the states of multi-agents are optimized based on the measurements from vehicular (e.g. VINS) and inter-agent measurements. Scholarly works on collaborative positioning are extensive. Any attempts to give a full relevant review of collaborative positioning would be incomplete. Based on how the states of the multi-agents are connected, we mainly divide the popular research streams of collaborative positioning into three groups: (1) the transponder-based, (2) the GNSS reference-based, and (3) the overlapping area detection-based.

The transponder-based collaborative positioning (CP) [22]–[24] make use of the radio ranging techniques, involving the time of arrival (TOA) [25], time difference of arrival (TDOA) [26] and received signal strength (RSS) [27], to estimate the relative distance, namely inter-ranging measurements, between multiple agents. Therefore, multi-agents are connected using the inter-ranging measurements and their states can be optimized by considering all the acquired measurements using weighted least squares [28] or factor graph optimization [29]. However, the accuracy of inter-ranging measurements is not guaranteed in urban canyon. Its performance can be significantly degraded due to the reflection or blockage of surroundings objects, such as buildings and dynamic vehicles. In short, the performance of inter-ranging measurements relies heavily on environmental conditions. To solve this problem, GNSS reference-based CP [28], [30] is proposed to derive the relative measurements between agents. The method does not require the inter-ranging and only low-level GNSS raw measurements are required. By double-differencing [28], [30] the GNSS measurements between two agents, the relative constraint can be derived to optimize the states of multiple agents. Significantly improved performance can be obtained in open areas. However, the accuracy of relative measurements estimated in [28], [30] suffers from the multipath effects and NLOS receptions in urban canyons. To solve this problem, our previous work in [31], [32] goes one step further by making use of the 3D building model to mitigate the impacts of multipath and NLOS signals before estimating the inter-agent measurements. Besides, a novel weighting scheme is proposed to model the uncertainty of the measurements from different agents. Improved performance is obtained in the tested dataset. However, the work in [31] relies heavily on the assumption that the master satellites are healthy without introducing any signal reflection or blockage. The pseudorange error involved in the signal transmitted from the selected master satellite will be delivered to the estimation of relative constraint. Unfortunately, it is hard to satisfy this assumption in urban canyons [33] with numerous tall buildings. Interestingly, the work in [21] proposes to make use of both the inter-ranging measurements and 3D mapping information to increase the robustness of the collaborative positioning. Improved performance is obtained with the help of the inter-ranging measurements. However, both of the work in [21], [31] share the drawbacks of the stringent reliance on the initial guess of vehicular states and the availability of 3D building models. Moreover, the accuracy of the [21], [31] is still far from the requirement of ITS which needs lane-level accuracy.

The other research stream [34]–[36] is the overlap detection-based collaborative positioning where the exteroceptive sensors, such as the 3D light detection and ranging (LiDAR) or camera, are employed to detect the overlapping area to further formulate the connection between agents. The work in [37] proposed a collaborative visual simultaneous localization and mapping (SLAM) framework to enhance the accuracy of state estimation of each robot equipped with a monocular camera. The connection between multiple robots is established based on loop closure (also called overlapping area) detection between agents. In other words, the connection between the two robots is built if the overlapping area is detected. Similar work in [38] explores the potential of collaborative visual SLAM using low-cost crowdsourcing data. The significantly improved positioning accuracy shows the advantage of collaborative positioning based on the onboard camera, compared with the single agent-based method. However, the robustness and accuracy are limited simply based on the monocular camera. The work in [36], [39] employs the IMU to help to estimate the scaling and perform VINS in each agent. A similar overlapping area detection scheme is employed to establish the inter-agent connections. Unfortunately, only relative measurements are derived from VINS or overlapping detection, causing inevitable drift in state estimation of agents.

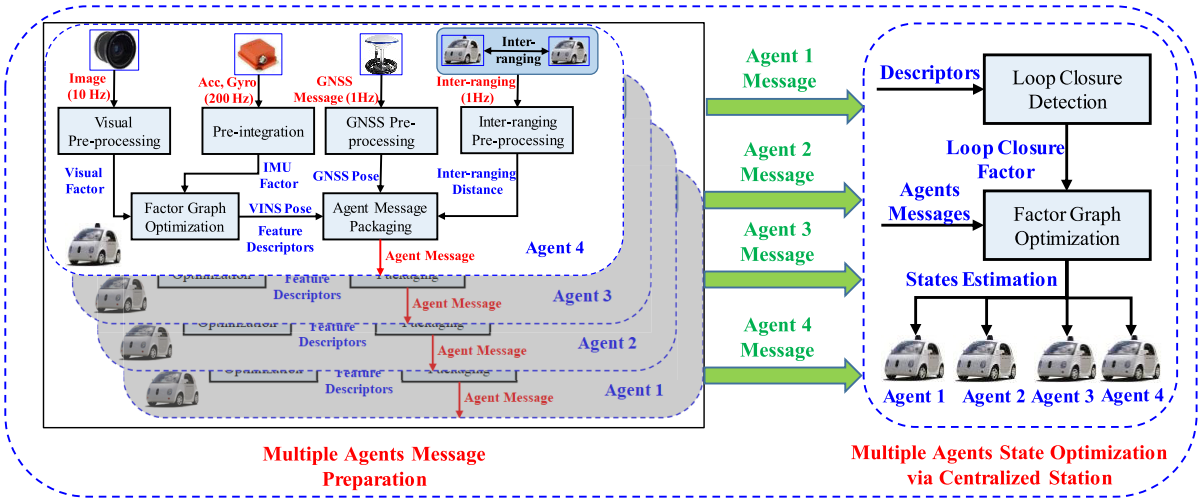


FIGURE 1. Overview of the proposed MCI framework which is composed of two parts: the multiple agent's message preparation, and multiple agents state optimization via centralized station.

Interestingly, the latest work in [34] extends the ultra-wideband (UWB) to help the collaborative positioning of the aerial swarm equipped with visual and inertial sensors. Each agent performs the VINS independently to derive the visual odometry relatively. Then the factor graph optimization is performed to estimate the relative pose between aerial swarms by integrating the measurements from VINS, UWB, and object tracking. However, only the relative pose between aerial swarms is estimated in indoor environments. Moreover, the collaborative framework is limited in indoor environments with reliance on UWB stations. In short, the research of overlap detection-based collaborative positioning is active in the robotics field and only relative positioning is achieved in indoor or constrained environments. Besides, the performance of VINS and camera-based overlapping area detection can be significantly impaired in urban canyons, due to the unexpected moving objects and unstable illumination. Its potential in outdoor urban environments is still to be explored.

Inspired by the work in [21] and [34], this paper proposes to leverage both the vision-based overlapping area detection and the inter-ranging measurements to boost the cross-connection between multi-agents and bring the MCI to outdoor urban scenarios using low-cost onboard sensors. The globally referenced GNSS solution is employed to mitigate the drift of the positioning of multiple agents. A novel MCI framework is proposed to integrate the sensor measurements from low-cost GNSS receiver, camera, IMU, and inter-ranging of each agent, using state-of-the-art factor graph optimization (FGO) to fully release the potential of MCI based on onboard low-cost sensors. To the best of the authors' knowledge, this is the first paper to employ both the overlapping area detection and inter-ranging to connect multiple agents to achieve the globally referenced collaborative positioning in urban canyons based on low-cost onboard sensors. We believe that the proposed framework can have a positive impact on both the academic and industrial fields.

II. OVERVIEW OF THE PROPOSED METHOD

The overview of the proposed MCI framework is shown in the following Fig. 1. The framework is composed of two parts: 1) the multiple agent's message preparation stage and 2) multiple agents state optimization via a centralized station.

Firstly, a single agent collects raw sensor measurements from multiple sensors and performs pre-processing steps to prepare an agent message, which will be sent to a centralized station for the optimization. The camera and IMU of a single agent form a minimal VINS sensor setup and the visual/inertial integration is performed using factor graph optimization to estimate the pose of the agent. The output of the integration is the VINS pose and feature descriptor. Secondly, The GNSS pre-processing block outputs the positioning information in the east, north, and up (ENU) coordinate system [12]. The inter-ranging preprocessing block outputs the inter-ranging distance between two separate agents on the road. Finally, the VINS pose, feature descriptors, GNSS positioning, and inter-ranging distance measurements are packaged to a customized agent message. In this paper, we assume that each agent is equipped with the same sensor setup and all the agent messages are sent to the centralized station via wireless communication. Moreover, the proposed MCI framework can also be applied to agents with different sensor setups.

In the centralized station side, all the agent messages are collected with a specific agent index. Then the loop closure detection is performed to detect the cross-connection between the separate agents based on the feature descriptors. Besides, the loop closure between different image frames of a single agent is also detected to mitigate the pose drift. Based on the received agent messages and newly detected loop closure detection results, a factor graph can be constructed based on the following constraints:

(1) The pose estimation from the VINS of each agent.

(2) The pose estimation from the GNSS receiver of each agent.
(3) The loop closure constraint of each agent.
(4) The overlapping area detection constraint between two separate agents.
(5) The inter-ranging measurements between two separate agents.

Therefore, the pose estimation of each agent can be re-optimized based on the newly established graph structure to obtain improved positioning accuracy. The major contributions of this paper are listed as follows:

(1) This paper proposes to leverage both the vision-based overlapping area detection and the inter-ranging measurements to boost the cross-connection between multi-agents and bring the MCI to outdoor challenging urban scenarios using low-cost onboard vehicular sensors.
(2) This paper proposes a novel MCI framework to integrate the sensor measurements from low-cost GNSS receiver, camera, IMU, and inter-ranging of multiple agents, using state-of-the-art factor graph optimization (FGO) to fully release the potential of MCI based on onboard low-cost sensors.
(3) This paper evaluates the performance of the proposed MCI using two challenging datasets collected in challenging urban canyons of Hong Kong.

To make the proposed MCI framework clearer, the following notations are defined and followed by the rest of the paper.

a) The VINS works in the local frame $\{L_i\}$ of the ith agent. The body frame of the ith agent is denoted by $\{B_i\}$.
b) The $\{\cdot\}^{L_i}$ denotes the variables concerning the ith agent's local frame.
c) The $\{\cdot\}^{b_i}$ denotes the variables concerning the ith agent's body frame.
d) The $\{\cdot\}^{w}$ denotes the variables concerning the world frame. The east north and up (ENU) coordinate is selected as the world frame in this paper.
e) The variable i denotes the index of the agent. The variable k denotes the index of the keyframe of a single agent.

III. AGENT MESSAGE PREPROCESSING

A. VISUAL/INERTIAL NAVIGATION SYSTEM (VINS)

For the visual/inertial integration, we adopt the existing VINS algorithm proposed in [15]. The VINS estimates the local pose of the agent. There are many impressive VINS algorithms, such as the [40]–[42]. Any of them can be applied to our proposed MCI framework. Although the VINS is not the contribution of this paper, we still briefly present the formulation of VINS optimization for the completeness. The state vector of the VINS for the ith agent in this paper is defined as follows,

$$\boldsymbol{\chi}_{\mathbf{i}} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \mathbf{x}_c^b, \lambda_1, \lambda_2, \ldots \lambda_M] \tag{1}$$

$$\mathbf{x}_k = [\mathbf{p}_{b_k}^{L_i}, \mathbf{v}_{b_k}^{L_i}, \mathbf{q}_{b_k}^{L_i}, \mathbf{b}_{a,k}, \mathbf{b}_{g,k}], \quad k \epsilon [0, n] \tag{2}$$

$$\mathbf{x}_c^b = [\mathbf{p}_c^b, \mathbf{q}_c^b] \tag{3}$$

where the superscript L_i denotes the local frame and subscript b_k denotes the body frame (the IMU frame) while taking the kth image. $\mathbf{x}_k$ is the IMU state at the corresponding to the kth image. It contains the position ($\mathbf{p}_{b_k}^{L_i}$), velocity ($\mathbf{v}_{b_k}^{L_i}$), and orientation represented by the quaternion ($\mathbf{q}_{b_k}^{L_i}$) in the world frame, and acceleration bias ($\mathbf{b}_{a,k}$) and gyroscope bias ($\mathbf{b}_{g,k}$) in the IMU body frame. n is the total number of keyframes considered for optimization and M is the total number of features considered. λ_l is the inverse depth of the lth feature observed for the first time, $l \in (1, M)$. $\mathbf{x}_c^b$ is the extrinsic parameter between the IMU frame and the camera frame. To guarantee the computation efficiency, we only make use of the measurements inside a sliding window to estimate the states. The images inside in the sliding window are between the frame b_k and b_{k+n}, with the time of t_k and t_{k+n}, respectively. The VINS estimation is formulated as a nonlinear least square problem for the ith agent as follows:

$$\begin{aligned} \underset{\boldsymbol{\chi}_i}{\operatorname{argmin}} \Big\{ & \left\| \mathbf{r}_p - \mathbf{H}_p \boldsymbol{\chi}_i \right\|^2 \\ & + \sum\nolimits_{k \epsilon \mathcal{B}} \left\| \mathbf{r}_{\mathcal{B}} \left(\hat{\mathbf{Z}}_{b_{k+1}}^{b_k}, \boldsymbol{\chi}_i \right) \right\|_{\mathbf{P}_{b_{k+1}}^{b_k}}^2 \\ & + \sum\nolimits_{(l,j) \epsilon \mathbf{C}} \left\| \mathrm{r}_C \left(\hat{\mathbf{Z}}_l^{c_j}, \boldsymbol{\chi}_i \right) \right\|_{\mathbf{P}_l^{c_j}}^2 \Big\} \end{aligned} \tag{4}$$

where $\{\mathbf{r}_p, \mathbf{H}_p\}$ is the prior information from the marginalization operation. The marginalization step is applied to marginalize the old states and measurements outside the sliding window. $\mathbf{r}_{\mathcal{B}}(.)$ is the residual term for IMU pre-integration. The pre-integration technique is employed to combine the high-frequency IMU measurements into a low-frequency IMU factor. The variable $\hat{\mathbf{Z}}_{b_{k+1}}^{b_k}$ denotes the pre-integration term which integrates the IMU measurements between keyframe b_k and b_{k+1}. The variable $\mathbf{P}_{b_{k+1}}^{b_k}$ denotes the covariance matrix corresponding to the $\hat{\mathbf{Z}}_{b_{k+1}}^{b_k}$. The variable $\mathbf{r}_C(.)$ is the residual term for visual re-projection of the feature. For a given new image, the features are detected by the Shi-Tomasi [43] corner detection algorithm. Meanwhile, the Kanade-Lucas-Tomasi (KLT) sparse optical flow algorithm [44] is employed to track the features. $\mathcal{B}$ is the set of all IMU measurements, $\mathbf{C}$ is the set of features that have been observed at least twice in the considered sliding window. $\mathbf{P}_l^{c_j}$ is the covariance matrix for visual re-projection, which represents the uncertainty of feature measurements. The Ceres-solver [45] is employed to solve the (4) to estimate the state set inside a sliding window. Regarding the implementation of the VINS in this paper, we refer to the framework in [15]. The detail of the VINS can be found in [15].

B. AGENT MESSAGE PACKAGING

For a given epoch k (corresponding to the $\mathbf{x}_k$) of an ith agent, the positioning information involves the $\mathbf{z}_{k,VINS}^{L_i} = [\mathbf{p}_{b_k}^{L_i}, \mathbf{q}_{b_k}^{L_i}, \mathbf{p}_c^b, \mathbf{q}_c^b]^T$ from VINS, GNSS positioning, and the

inter-ranging (see Fig. 1). The reference frame of VINS is the first frame of the IMU, namely where the agent starts.

The GNSS receiver provides the positioning in the earth-centered, earth-fixed (ECEF) coordinate system [12]. In this paper, we convert the ECEF coordinate into the ENU frame to denote the GNSS positioning by setting the first GNSS measurements of the first agent as the reference point of the ENU frame. For a given epoch k of an ith agent, the GNSS positioning in ENU is given as $\mathbf{z}_{k,GNSS}^{w,i} = [x_k^{w,i}, y_k^{w,i}, z_k^{w,i}]^T$. The subscript w denotes the original of the ENU frame, namely the world frame. The inter-ranging measurement providing the distance between two separate agents i and j, denoted by $z_{k,IR}^{i,j}$. The subscript IR denotes the inter-ranging measurement.

To detect the overlapping area between agents, the visual information is needed to be transmitted to the centralized station. The straightforward solution is to deliver the keyframe image b_k. However, this will lead to an unacceptable message transmission load. Considering the requirements of overlapping detection, we only deliver the feature descriptors and 3D position of landmark which are estimated by the VINS. The Binary robust independent elementary features (BRIEF) are employed to describe the features detected using the Shi-Tomasi corner detection algorithm. The set of BRIEF is denoted by $\mathbf{F}_k^{BRIEF} = [f_{k,1}^{BRIEF}, f_{k,2}^{BRIEF}, \ldots]^T$ at a given epoch k. The position landmark set is given by $\mathbf{P}_k^{BRIEF} = [p_{k,1}^{BRIEF,L_i}, p_{k,2}^{BRIEF,L_i}, \ldots]^T$ in the L_i frame.

To sum up, the following information is packaged as an agent keyframe message (AKM) as follows:

(1) Local pose estimation from VINS: $\mathbf{z}_{k,VINS}^{L_i} = [\mathbf{p}_{b_k}^{L_i}, \mathbf{q}_{b_k}^{L_i}]^T$ involving the positioning and orientation relative to the local frame of VINS.
(2) GNSS positioning in the world frame as $\mathbf{z}_{k,GNSS}^{w,i} = [x_k^{w,i}, y_k^{w,i}, z_k^{w,i}]^T$.
(3) The inter-ranging measurement between two separate agents i and j, denoted by $z_{k,IR}^{i,j}$.
(4) Feature descriptors $\mathbf{F}_k^{BRIEF} = [f_{k,1}^{BRIEF}, f_{k,2}^{BRIEF}, \ldots]^T$ and corresponding 3D position $\mathbf{P}_k^{BRIEF} = [p_{k,1}^{BRIEF,L_i}, p_{k,2}^{BRIEF,L_i}, \ldots]^T$ which is to be used for overlapping area detection.

As we do not need to send the raw image and only the processed measurements are transmitted to the centralized station, the bandwidth of data transmission is limited to a low bound.

IV. MULTI-AGENTS COLLABORATIVE INTEGRATION (MCI)

The centralized station receives the AKM from multiple agents. First, the overlapping area detection is performed to detect the inter-agent connections based on the feature descriptors and 3D landmarks from AKM. Meanwhile, the loop closure of a single agent is also detected using the same algorithms. The outputs of both the overlapping area detection and the loop closure detection are the relative pose

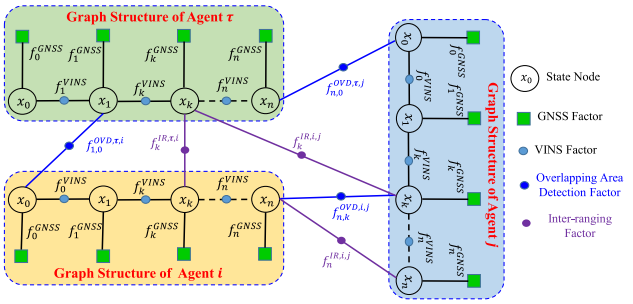


FIGURE 2. Factor graph for multi-agents (τth, ith, and jth agents) collaborative integration.

estimation between two AKMs. Second, the measurements from AKM include two parts, the local measurements relative to the local frame (e.g. pose estimation arising from VINS), or the global measurements relative to the global frame (e.g. GNSS positioning). The relative measurements are subjected to drift over time. Meanwhile, the global measurements are noisy but drift-free.

Based on all the messages received from multiple agents and overlapping detection results, a factor graph [29] can be established as Fig. 2. The state node involves the orientation and translation of each AKM relative the world frame and is denoted as $\mathbf{x}_k^{w,i} = [\mathbf{p}_{b_k}^{w,i}, \mathbf{q}_{b_k}^{w,i}]^T$. The subscripts i and k denote the agent ID and index of the AKM, respectively. Therefore, the state set of the global optimization is as follows:

$$\boldsymbol{\chi}_{MCI} = [\mathbf{x}_0^{w,0}, \mathbf{x}_1^{w,0}, \ldots, \mathbf{x}_0^{w,1}, \mathbf{x}_1^{w,1}, \ldots, \mathbf{x}_0^{w,i}, \mathbf{x}_1^{w,i}, \ldots] \quad (5)$$

the $\mathbf{x}_0^{w,0}, \mathbf{x}_1^{w,0}, \ldots$ denotes the state set of agent 0. The $\mathbf{x}_0^{w,1}, \mathbf{x}_1^{w,1}, \ldots$ denotes the state set of agent 1. The $\mathbf{x}_0^{w,i}, \mathbf{x}_1^{w,i}, \ldots$ denotes the state set of agent i.

The three shaded areas (green, yellow, and blue rectangles in Fig. 2) denote the factor graph for agents τ, i, and j, respectively. Within the factor graph of a single agent, the light blue circle and small green rectangles denote the VINS and GNSS factors, respectively. The purple and blue circles connecting two separate agents represent the overlapping area detection and the inter-ranging factors, respectively. The factor graph optimization is performed to optimize all the state nodes inside the factor graph of each agent with the help of the inter-agent connections, the overlapping area detection, and the inter-ranging factor. In other words, optimizing the factor graph in Fig. 2 equals to find a configuration of nodes that matches all the factors as much as possible.

To make the global optimization clear, Fig. 3 shows the coordinate transformations of the ith and jth agents. The $T_{L_i}^W$ encodes the transformation from the local frame (L_i) of ith agent to the world frame. The $T_{B_i}^{L_i}$ encodes the transformation from the body frame (B_i) to the local frame (L_i) of ith agent. Be noted that the $T_{B_i}^{L_i}$ is a homogeneous transformation that involves both the rotation and translation. The transformation is also the same for the jth agent which can be seen in Fig. 3. The local measurements, such as the VINS,

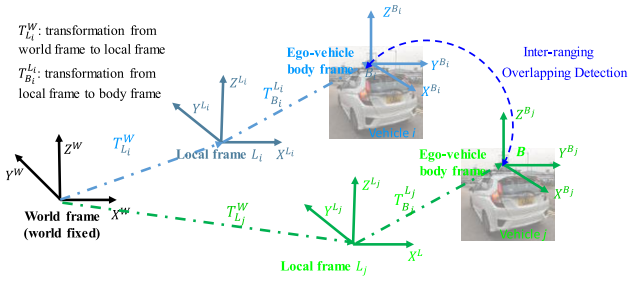


FIGURE 3. Coordinate transformations of *i*th and *j*th agents. The blue dash curves denote the inter-agents connection established by inter-ranging and overlapping detection.

provide the constraint in the local frame as $T_{B_i}^{L_i}$. The global measurements, such as the GNSS positioning, provide the constraint in the world frame as $T_{B_i}^{W}$. Optimizing the factor graph in Fig. 2 equals to finding the optimized $T_{L_i}^{W}$ of each agent that matches all the local and global measurements. To guarantee real-time performance, global optimization is performed at a frequency of 1 Hz in this paper. However, the $T_{B_i}^{L_i}$ is given at a frequency of VINS. Therefore, the pose estimation of each agent relative to the world frame can be derived as $T_{L_i}^{W} T_{B_i}^{L_i}$ at a frequency of VINS with real-time performance.

The rest of Section IV presents the derivation of the listed factors above and corresponding optimization.

A. VINS FACTOR

The VINS can provide accurate and relative pose estimation in a short period, this paper makes use of the VINS to constraint the pose of the agent between two frames. As the pose estimation from the VINS is given as $\mathbf{z}_{k,VINS}^{L_i} = [\mathbf{p}_{b_k}^{L_i}, \mathbf{q}_{b_k}^{L_i}]^T$, the error function ($\mathbf{e}_{k,VINS}^{i}$) for the VINS factor can be derived as follows:

$$\begin{aligned}\mathbf{e}_{k,VINS}^{i} &= \left(\mathbf{z}_{k,VINS}^{L_i} \ominus \mathbf{z}_{k-1,VINS}^{L_i}\right) - h_{k,VINS}^{i}\left(\mathbf{x}_{k}^{w,i}, \mathbf{x}_{k-1}^{w,i}\right) \\ &= \begin{bmatrix} \mathbf{q}_{b_{k-1}}^{L_i-1}\left(\mathbf{p}_{b_k}^{L_i} - \mathbf{p}_{b_{k-1}}^{L_i}\right) \\ \mathbf{q}_{b_{k-1}}^{L_i-1}\mathbf{q}_{b_k}^{L_i} \end{bmatrix} \\ &\quad \ominus \begin{bmatrix} \mathbf{q}_{w}^{L_i}\mathbf{q}_{b_{k-1}}^{w-1}\left(\mathbf{p}_{b_k}^{w} - \mathbf{p}_{b_{k-1}}^{w}\right) \\ \mathbf{q}_{w}^{L_i}\mathbf{q}_{b_{k-1}}^{w-1}\mathbf{q}_{b_k}^{w} \end{bmatrix} \end{aligned} \tag{6}$$

where the variables $\mathbf{x}_{k}^{w,i}$ and $\mathbf{x}_{k-1}^{w,i}$ denotes the states of the *i*th agent at epoch *k* and *k-1*, respectively. The observation $h_{k,VINS}^{i}(\cdot)$ connects the state ($\mathbf{x}_{k}^{w,i}$) and observation measurement ($\mathbf{z}_{k,VINS}^{L_i}$) from VINS. The symbol $\ominus$ is the minus operation on the error state of a quaternion. The first row of $\mathbf{e}_{k,VINS}^{i}$ denotes the relative pose error between epoch *k* and *k-1*. The second row denotes the relative rotation error. The covariance matrix of the $\mathbf{e}_{k,VINS}^{i}$ is given as $\mathbf{\Sigma}_{k,VINS}^{i}$ and is tuned based on the intrinsic parameters of the camera. In this case, each VINS pose estimation can provide a constraint to the factor graph in Fig. 2.

FIGURE 4. Illustration of overlapping area between agent 1 and 3 near an intersection. The red circles denote the detected features using the Shi [43] corner detection algorithm.

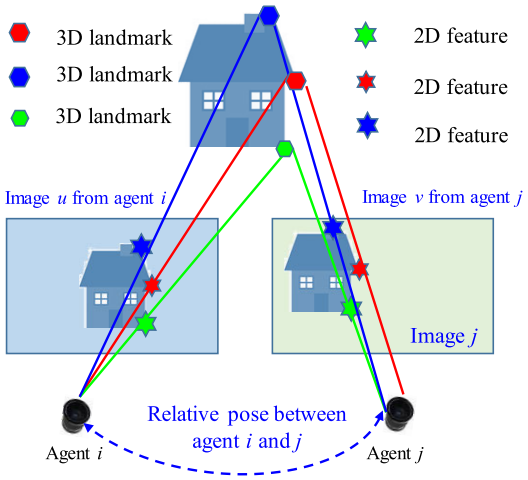


FIGURE 5. Illustration of using the PnP algorithm to estimate relative pose between two frames.

B. OVERLAPPING AREA DETECTION FACTOR

Overlapping area detection is an effective way to establish the connection between agents as the blue factors shown in Fig. 2. Fig. 4 shows two images from two separate agents passing the same intersection. In this paper, we make use of the method of loop closure detection in visual SLAM to detect the overlapping based on the images from two agents. Then the relative pose between two agents is estimated using the perspective-n-Point (PnP) algorithm [46].

Firstly, a bag-of-words (DBoW2) [47] place recognition method is employed to detect the overlapping between two images (see Fig. 5), the image *u* from agent *i*, and image *v* from agent *j*. DBoW2 returns the matching scores of two

images based on the BRIEF descriptors. We argue that the overlapping between two agents i and j is detected when the matching score is beyond a threshold. Then the relative pose, denoted by $\mathrm{z}^{i,j}_{u,v,OVD} = [\mathbf{p}^{i,j}_{OVD}, \mathbf{q}^{i,j}_{OVD}]^T$, between two keyframes u and v can be estimated using PnP. The following Fig. 5 shows the PnP problem based on two images from agent i and j. The colored stars show the detected 2D features. The colored pentagons denote the 3D landmarks whose positions are estimated by VINS in Section III-A. The inputs of the PnP algorithm involve the 3D pose of the landmarks and the pixel positions of the 2D features.

Therefore, the error function ($\mathbf{e}^{i,j}_{u,v,OVD}$) for the overlapping area detection factor can be derived as follows:

$$\begin{aligned}\mathbf{e}^{i,j}_{u,v,OVD} &= \left(\mathrm{z}^{i,j}_{u,v,OVD}\right) - h^{i,j}_{u,v,OVD}\left(\mathbf{x}^{w,i}_{u}, \mathbf{x}^{w,j}_{v}\right) \\ &= \begin{bmatrix}\mathbf{p}^{i,j}_{OVD} \\ \mathbf{q}^{i,j}_{OVD}\end{bmatrix} \ominus \begin{bmatrix}\mathbf{q}^{w,i-1}_{u}\left(\mathbf{p}^{w,j}_{v} - \mathbf{p}^{w,i}_{u}\right) \\ \mathbf{q}^{w,i-1}_{v}\mathbf{q}^{w,j}_{v}\end{bmatrix}\end{aligned} \tag{7}$$

the variable $\mathbf{x}^{w,i}_{u}$ denote the pose of the keyframe image u from agent i in the world frame. Similarly, the $\mathbf{x}^{w,j}_{v}$ denote the pose of the keyframe image v from agent j in the world frame. The variables $\mathbf{p}^{i,j}_{OVD}$ and $\mathbf{q}^{i,j}_{OVD}$ denote the position and rotation of the relative pose between keyframes u and v. The $h^{i,j}_{u,v,OVD}(\cdot)$ denotes the observation function connects the states and the $\mathrm{z}^{i,j}_{u,v,OVD}$. The covariance matrix of $\mathbf{e}^{i,j}_{u,v,OVD}$ is given as $\mathbf{\Sigma}^{i,j}_{u,v,OVD}$ and is tuned experimentally with fixed values. In this case, each $\mathbf{e}^{i,j}_{u,v,OVD}$ provides a constraint, denoted by the blue circle, to the factor graph in Fig. 2.

C. GNSS FACTOR

The GNSS receiver provides absolute positioning given by $\mathbf{z}^{w,i}_{k,GNSS} = [x^{w,i}_{k}, y^{w,i}_{k}, z^{w,i}_{k}]^T$ from the ith agent at epoch k in the world frame. The GNSS positioning can help to mitigate the overall drift in long-term driving. The error function ($\mathbf{e}^{i}_{k,GNSS}$) for the measurement $\mathbf{z}^{w,i}_{k,GNSS}$ can be given as follows:

$$\begin{aligned}\mathbf{e}^{i}_{k,GNSS} &= \mathbf{z}^{w,i}_{k,GNSS} - h^{i}_{k,GNSS}\left(\mathbf{x}^{w,i}_{k}\right) \\ &= \begin{bmatrix}x^{w,i}_{k} \\ y^{w,i}_{k} \\ z^{w,i}_{k}\end{bmatrix} - \begin{bmatrix}(\mathbf{p}^{w,i}_{b_k})_x \\ (\mathbf{p}^{w,i}_{b_k})_y \\ (\mathbf{p}^{w,i}_{b_k})_z\end{bmatrix}\end{aligned} \tag{8}$$

the $h^{i}_{k,GNSS}$ denotes the observation function which connects the $\mathbf{z}^{w,i}_{k,GNSS}$ and state $\mathbf{p}^{w,i}_{b_k}$ of the ith agent at epoch k. The operator $(\mathbf{p}^{w,i}_{b_k})_*$ is employed to get a certain component of $\mathbf{p}^{w,i}_{b_k}$. The covariance matrix of $\mathbf{e}^{i}_{k,GNSS}$ is given as $\mathbf{\Sigma}^{i}_{k,GNSS}$ and is directly given from the output of the GNSS receiver.

D. INTER-RANGING FACTOR

The inter-ranging measurements provide the relative distance between agents as shown in Fig. 6. The variable $\mathrm{z}^{i,j}_{k,IR}$ denotes the distance between agent i and j. For the given inter-ranging connections shown in Fig. 6, the error function can be derived

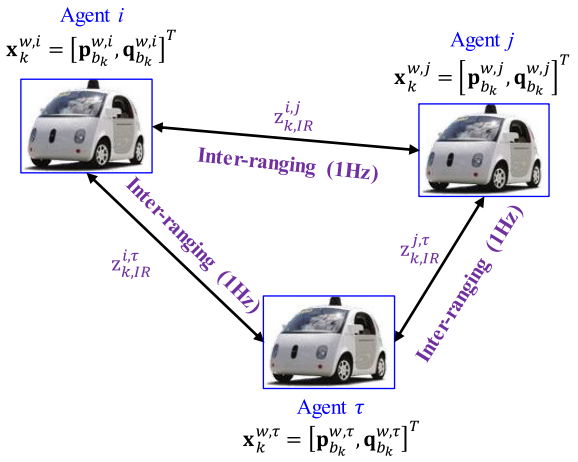


FIGURE 6. Illustration of inter-ranging measurements between three agents. The model of the small car is from Waymo [48].

as follows:

$$\mathbf{e}^{i,j}_{k,IR} = \mathrm{z}^{i,j}_{k,IR} - h^{i,j}_{k,IR}\left(\mathbf{x}^{w,i}_{k}, \mathbf{x}^{w,j}_{k}\right) = \mathrm{z}^{i,j}_{k,IR} - \left\|\mathbf{x}^{w,i}_{k} - \mathbf{x}^{w,j}_{k}\right\| \tag{9}$$

the $\left\|\mathbf{x}^{w,i}_{k} - \mathbf{x}^{w,j}_{k}\right\|$ denotes the Euclidean distance between agent i and j. The function $h^{i,j}_{k,IR}(\cdot)$ denotes the observation function which connects the observation and states. The covariance matrix of $\mathbf{e}^{i,j}_{k,IR}$ is given as $\mathbf{\Sigma}^{i,j}_{k,IR}$ and is experimentally determined.

Similarly, the error functions can also be derived as follows:

$$\mathbf{e}^{i,\tau}_{k,IR} = \mathrm{z}^{i,\tau}_{k,IR} - h^{i,\tau}_{k,IR}\left(\mathbf{x}^{w,i}_{k}, \mathbf{x}^{w,\tau}_{k}\right) = \mathrm{z}^{i,\tau}_{k,IR} - \left\|\mathbf{x}^{w,i}_{k} - \mathbf{x}^{w,\tau}_{k}\right\| \tag{10}$$

$$\mathbf{e}^{j,\tau}_{k,IR} = \mathrm{z}^{j,\tau}_{k,IR} - h^{j,\tau}_{k,IR}\left(\mathbf{x}^{w,j}_{k}, \mathbf{x}^{w,\tau}_{k}\right) = \mathrm{z}^{j,\tau}_{k,IR} - \left\|\mathbf{x}^{w,j}_{k} - \mathbf{x}^{w,\tau}_{k}\right\| \tag{11}$$

The covariance matrix for the inter-ranging measurements is given with the same value.

E. OPTIMIZATION-BASED MULTI-AGENTS INTEGRATION

Based on the listed four kinds of factors, the optimal estimation of the $\boldsymbol{\chi}_{MCI}$ can be formulated as the following nonlinear least square (NLS) problem,

$$\begin{aligned}\boldsymbol{\chi}_{MCI}{}^{*} = \underset{\boldsymbol{\chi}_{MCI}}{\operatorname{argmin}} &\sum\nolimits_{i=1}^{I}\sum\nolimits_{k\in V_i}\left\|\mathbf{e}^{i}_{k,VINS}\right\|^2_{\mathbf{\Sigma}^{i}_{k,VINS}} \\ &+ \sum\nolimits_{(i,j,u,v\in OVD_{i,j})}\left\|\mathbf{e}^{i,j}_{u,v,OVD}\right\|^2_{\mathbf{\Sigma}^{i,j}_{u,v,OVD}} \\ &+ \sum\nolimits_{(i,j,k\in IR_{i,j,k})}\left\|\mathbf{e}^{i,j}_{k,IR}\right\|^2_{\mathbf{\Sigma}^{i,j}_{k,IR}} \\ &+ \sum\nolimits_{i=1}^{I}\sum\nolimits_{k\in G_i}\left\|\mathbf{e}^{i}_{k,GNSS}\right\|^2_{\mathbf{\Sigma}^{i}_{k,GNSS}}\end{aligned} \tag{12}$$

the variable $\boldsymbol{\chi}_{MCI}{}^{*}$ denotes the optimal estimation of $\boldsymbol{\chi}_{MCI}$.The variable V_i denotes the set of VINS factors. The variable $OVD_{i,j}$ denotes the set of overlapping area detection

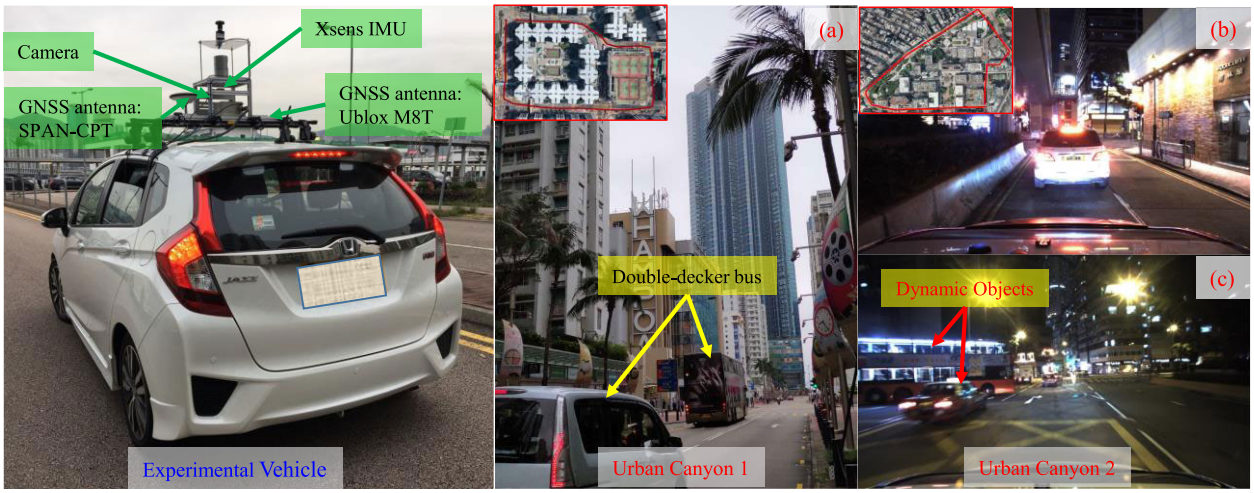


FIGURE 7. Overview of the sensor setups and scenarios for evaluations. (a): scene of the evaluated urban canyon 1. (b) and (c): scene of the evaluated urban canyon 2.

factors. The variable $IR_{i,j,k}$ denotes the set of inter-ranging factors. The variable G_i denotes the set of GNSS factors. Similar to the previous definition, the variables i and j denote the index of agents. The variable k represents the index of epoch inside agents. The variable I denotes the total number of agents considered in MCI.

The Ceres-solver [45] is employed to solve the optimization problem using the Levenberg-Marquardt method. Once the optimization is solved, the $T_{L_i}^W$ can be calculated as follows:

$$T_{L_i}^W = T_{B_i}^W T_{B_i}^{L_i - 1} \tag{13}$$

the variable $T_{B_i}^{L_i}$ is given by the local pose estimation, the VINS. In the implementation, the optimization is performed once a new GNSS factor is added to the factor graph in Fig. 2.

V. EXPERIMENT RESULTS

The proposed MCI method is verified through two real road datasets collected in the urban canyons of Hong Kong. The first experiment is implemented in a typical urban canyon of Hong Kong which can be seen in Fig. 7-(a). The evaluated trajectory is shown in the top-left side of Fig. 7-(a). The second experiment is conducted in a more complex urban canyon with more moving objects and high-rising buildings which can be seen in Fig. 7-(b∼c), leading to more challenges to the proposed MCI framework. The evaluated trajectory is shown in the top-left side of Fig. 7-(b). During the experiment, we drive several times with different routes to simulate multi-agents due to the limited experimental setup. We are fully aware that the conducted validation cannot fully replicate the time delay of message transmission between agents and the centralized station. The experimental vehicle is shown on the left side of Fig. 7. The coordinate systems between all the sensors were calibrated before the experiment. Then all the data are recorded using the robot operation system (ROS) [49] and then performed the proposed MCI framework offline using a centralized station.

Regarding the experiment is urban canyon 1, a monocular camera is employed to collect the raw color image at a frequency of 10 Hz. An Xsens MTi 30 IMU is employed to collect raw measurements at a frequency of 200 Hz. Besides, the NovAtel SPAN-CPT, a dual-frequency GNSS (GPS, GLONASS, and Beidou) RTK/INS (fiber-optic gyroscopes, FOG) integrated navigation system, was used to provide the ground truth of positioning evaluation. The gyro bias in-run stability of the FOG is 1 degree per hour and its random walk is 0.067 degree per hour. The baseline between the rover and GNSS base stations is less than 7 km. Regarding the inter-ranging measurements, we follow the work in [21] to simulate the inter-ranging measurements with Gaussian noise of $\mathcal{N}_{IR} = (0, 1.5m^2)$ based on the ground truth positioning from NovAtel SPAN-CPT. Meanwhile, the GNSS solution is also simulated with Gaussian noise of $\mathcal{N}_{IR} = (0, 10m^2)$ based on the ground truth positioning from NovAtel SPAN-CPT. We want to see how the proposed method can work when the GNSS noise is subject to the Gaussian assumption. The standard deviation of 10 meters for the simulated GNSS positioning is selected based on our previous experimental validation [11] in the urban canyon 1.

In experimental 2, we go one step further by validating the proposed MCI framework using a more challenging dataset. Different from the experimental 1, the GNSS positioning is provided using a commercial-level u-blox M8T GNSS receiver. The error distribution of the GNSS solution is highly non-Gaussian due to the unexpected signal reflection and blockage arising from tall buildings in urban canyon 2. Moreover, the density of the dynamic objects is significantly denser which can significantly challenge the performance of VINS, compared with the scene in experiment 1. The same simulation method is applied to simulate inter-ranging measurements.

TABLE 1. Positioning performances of different agents using different methods in urban canyon 1.

Methods		Agent 1		Agent 2		Agent 3		Agent 4	
		MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD
Single Agent Positioning	GNSS	12.49 m	7.09 m	11.65 m	6.80 m	13.07 m	6.77	12.07 m	6.37 m
	VINS (RPE)	0.625 m	0.522 m	0.49 m	0.36 m	0.45 m	0.64 m	0.87 m	0.92 m
	VINS (APE)	16.81 m	6.76 m	7.99 m	2.40 m	23.16 m	12.58 m	21.59 m	8.62 m
	GNSS/VINS	5.49 m	3.33 m	6.13 m	3.92 m	5.01 m	3.45 m	5.71 m	3.85 m
Multi-Agents Positioning	GNSS/VINS/OVD	4.22 m	1.95 m	5.33 m	1.70 m	4.39 m	4.86 m	3.94 m	2.01 m
	GNSS/VINS/OVD /IR	3.81 m	1.82 m	3.66 m	1.16 m	3.42 m	1.58 m	3.63 m	1.82 m

To verify the effectiveness of the proposed method, several methods are compared.

(1) **GNSS**: GNSS standalone positioning provided by the simulation in experiment 1 and the u-blox M8T receiver in experiment 2 for single-agent positioning.
(2) **VINS** [15]: VINS solution based on the work in [15] for single-agent positioning.
(3) **GNSS/VINS**: GNSS/VINS integration using FGO for single-agent positioning.
(4) **GNSS/VINS/OVD**: Multi-agents collaborative positioning by integrating the constraints from GNSS, VINS, overlapping area detection between agents, and loop closure inside agents.
(5) **GNSS/VINS/OVD/IR**: Multi-agents collaborative positioning by integrating the constraints from GNSS, VINS, overlapping area detection between agents, loop closure inside agents, and the inter-ranging measurements.

Regarding the performance evaluation, the popular EVO toolkit [50] is employed to calculate the absolute positioning errors (APE) of the listed five methods above. As only relative positioning is provided from VINS, we also evaluate the relative positioning error (RPE) for the VINS of a single agent.

A. EXPERIMENTAL EVALUATION IN URBAN CANYON 1

The positioning performance of the proposed method in urban canyon 1 is shown in Table 1. In this experiment, data from 4 agents are integrated using the proposed MCI framework. Both the single-agent positioning and multi-agents positioning solutions are compared. An APE of 12.49 meters is obtained based on GNSS standalone positioning for agent 1 with a standard deviation of 7.09 meters. Similar performances are obtained for the other agents regarding the GNSS standalone positioning, as similar Gaussian noise is employed to simulate the GNSS solutions. The fourth row of Table 1 shows the relative positioning error of VINS. The relative positioning error is significantly smaller than the APE since the RPE only accumulate in a short period between two epochs. An RPE of 0.625 meters is achieved based on VINS for agent 1. Interestingly, both the RPEs of agent 2 and 3 are slightly smaller than the one for agent 1, which is mainly due to the diverse environmental conditions. 0.87 meters of RPE is obtained for agent 4 with a standard deviation (STD) of 0.92 meters. Therefore, we can see that the VINS can provide decent positioning relatively in a short period. In the SLAM field, the research tends to make use of the RPE to denote the performance of VINS as only relative positioning is derived from the VINS. However, we still present the APE of the VINS for completeness. The fifth row of Table 1 shows the APE of VINS, which denotes the error accumulation. An APE of 16.81 meters is obtained from VINS for agent 1 with an STD of 6.76 meters. Interestingly, we can see that the APE is even larger than the APE of GNSS standalone positioning (12.49 meters). Moreover, both the APEs for agent 3 and 4 reaches more than 20 meters. In short, we conclude that the VINS can provide decent positioning in a short period and the GNSS is drift-free. Inspired by this, we also show the performance of the integration of VINS and GNSS in the sixth row of Table 1. Interestingly, the accuracies are significantly improved compared with both the VINS or GNSS standalone positioning for all the agents. An APE of 5.49 meters is obtained for the agent 1 with an STD of 3.33 meters. The APE is significantly smaller than both the GNSS (12.49 meters) and VINS (16.81 meters) standalone positioning. A similar phenomenon can also be seen by other agents. The results show that the positioning from the VINS and GNSS are complementary. However, the remaining APE still reaches more than 5 meters for all the agents using GNSS/VINS integration.

With the help of the overlapping area detection constraint using the proposed MCI framework, the APE is decreased from 5.49 meters to 4.22 meters. More importantly, the STD is also significantly reduced from 3.33 to 1.95 meters. Similar improvements can also be seen in the other agents. The improved accuracy shows that the proposed method which involves the OVD can effectively help to improve the positioning performance of all the agents. The main reason behind the improvement is that the OVD constraints connects all the states of the agents and corrects the corresponding potential errors by solving (12). However, we can see that the improvement of the OVD constraint for agent 3 is limited with the APE being improved from 5.01 meters to about 4.4 meters. One of the major reasons for this is the misdetection of OVD constraints. As we can see from Fig. 7, numerous dynamic

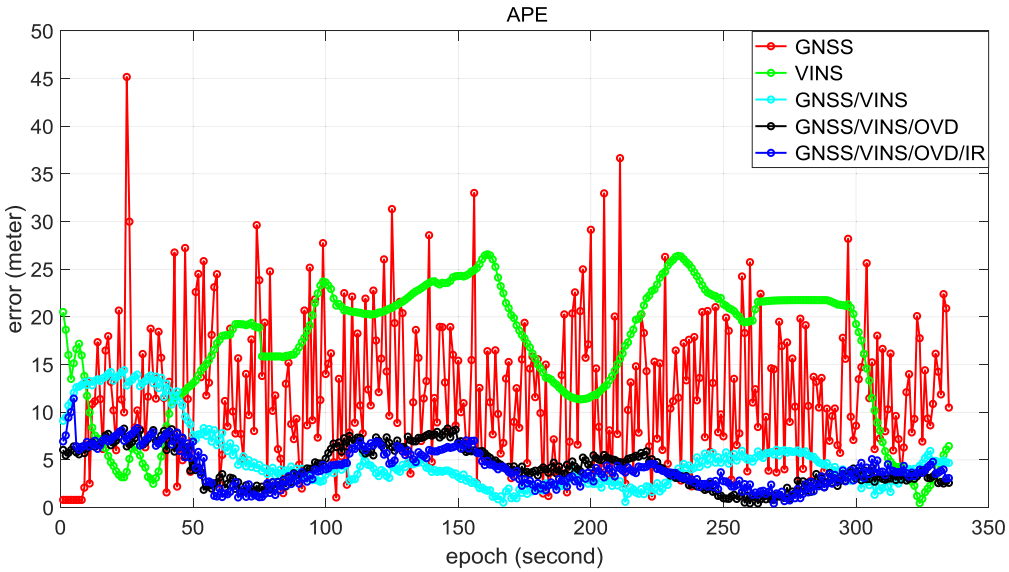


FIGURE 8. Illustration of the performance of the listed five methods for agent 1 during the evaluated dataset collected in urban canyon 1. The x-axis denotes the epoch during the test. The y-axis denotes the APE.

objects pose additional challenges to the applied OVD detection algorithm presented in Section IV-B. How to effectively identify the misdetection case is still an open question to see. Fortunately, with the help of the applied inter-ranging measurements, the APE is significantly improved for almost all the agents. The APE is decreased from 5.33 to 3.66 meters for agent 2. Meanwhile, the STD is also decreased slightly. A similar improvement can also be seen for agent 3. In short, the inter-ranging measurement can effectively improve the overall performance of all the agents, which shows the effectiveness of the proposed method.

To show the detail of the performance of a single agent using the listed five methods during the evaluation, we present the APEs of agent 1 in Fig. 8. The red and green curves denote the APE of GNSS and VINS standalone positioning, respectively. The GNSS positioning fluctuates dramatically during the test. The APE of VINS is significantly smoother compared with the GNSS. However, the maximum error of VINS reaches more than 25 meters which is far from the positioning requirement of autonomous driving. The cyan curve denotes the APE of GNSS/VINS integration with significantly smaller errors. However, the maximum error can still reach more than 10 meters. The black and blue curves denote the APE of both the GNSS/VINS/OVD and the GNSS/VINS/OVD/IR integrations, respectively. The performance is improved almost throughout the evaluated dataset, which shows the effectiveness of the proposed MCI framework. Fig. 9 shows the trajectories of the listed five methods for agent 1. The x-axis denotes the north direction during the test. The y-axis denotes the east direction. The red curve denotes the trajectory of GNSS standalone positioning. The green curve denotes the VINS solution. The cyan curve represents the trajectory of GNSS/VINS integration. The black and blue curves denote the trajectories based on proposed MCI frameworks, respectively. The magenta curve

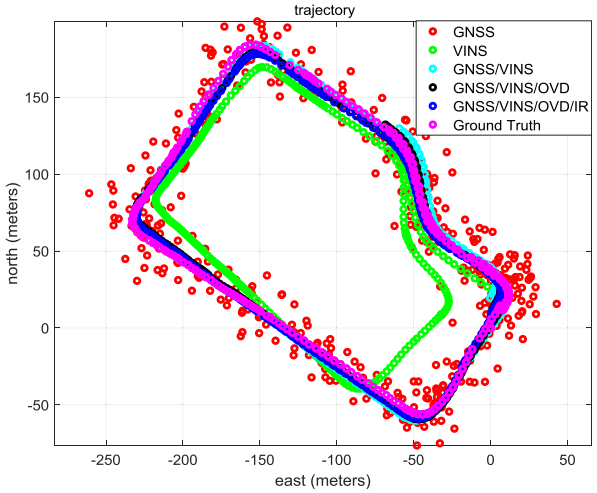


FIGURE 9. Illustration of the trajectories of the listed five methods for agent 1 during the evaluated dataset collected in urban canyon 1. The x-axis denotes the north direction during the test. The y-axis denotes the east direction. The red curve denotes the trajectory of GNSS standalone positioning. The green curve denotes the VINS solution. The cyan curve represents the trajectory of GNSS/VINS integration. The black and blue curves denote the trajectories based on proposed MCI frameworks, respectively. The magenta curve denotes the reference trajectory from SPAN-CPT.

denotes the reference trajectory from SPAN-CPT. We can see that the trajectory of the VINS (green curve) deviates significantly from the reference trajectory. The GNSS positioning fluctuates dramatically from the reference trajectory (magenta curve). The proposed MCI framework achieved smooth and accurate trajectories concerning the reference trajectory, which again shows the effectiveness of the proposed MCI framework.

Interestingly, the APE of GNSS/VINS integration is even smaller than the proposed MCI framework at epoch 150. The main reason behind it is the misdetection of the OVD

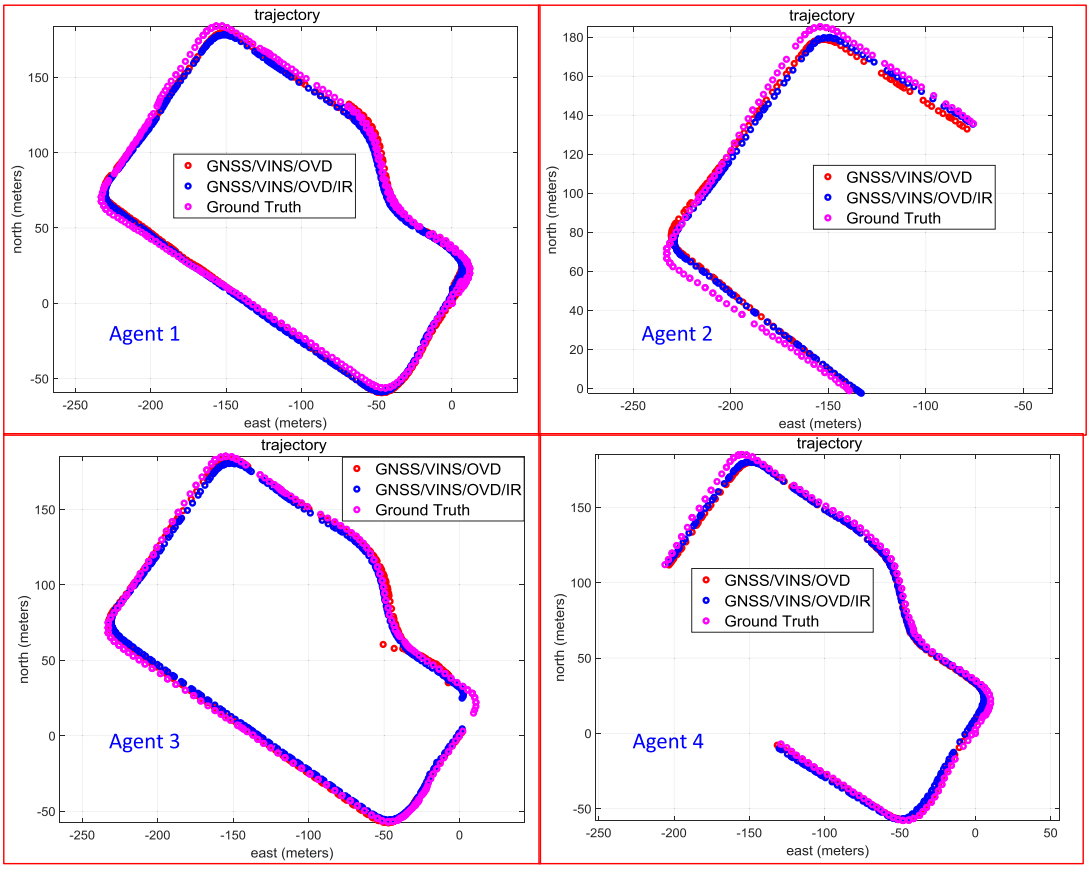


FIGURE 10. Illustration of the trajectories of four agents using the proposed MCI frameworks in urban canyon 1.

constraint. the misdetection of OVD can introduce a large positioning error to the whole factor graph shown in Fig. 2. One possible solution to mitigate the number of misdetection of OVD is properly tuning the threshold of the degree of feature matching [47]. However, the threshold is hard to adapt to different scenarios. Meanwhile, the error can also be contributed by the inter-ranging measurements. With the significantly increased amount of measurements from multiple agents, effectively identifying the fault measurements is significant for improving the overall positioning performance of all the agents. The switchable constraint is proposed in [51] to model the unexpected outlier measurements. However, it relies heavily on the data redundancy of healthy measurements. It is interesting to see how the switchable constraint can work in our proposed MCI framework which possesses a large amount of data redundancy.

Fig. 10 shows the trajectories of four agents using the proposed MCI frameworks (both the GNSS/VINS/OVD and GNSS/VINS/OVD/IR integrations). The top left of Fig. 10 shows the trajectories of agent 1. The red curve denotes the trajectory of agent 1 using GNSS/VINS/OVD integration. the blue and magenta curves denote the trajectories of GNSS/VINS/OVD integration and ground truth, respectively. The color representations are also followed for the other three agents shown in Fig. 10. We can see that the agent 1 possess the longest trajectory with the overlapping areas with all the other three agents.

With more agents participating in the MCI framework, more overlapping constraints are added to the optimization of (12). Inspired by this, we show the performance of the proposed MCI framework when fewer agents are considered in the MCI framework. Table 2 shows the performance of the MCI with only two agents (agents 1 and 2) being considered. An APE of 5.02 meters is obtained using the GNSS/VINS/OVD integration, which is larger than the one from Table 1 (4.22 meters). Similar performance improvement can also be seen for the GNSS/VINS/OVD/IR integration with the APE decreasing from 4.03 to 3.81 meters. Table 3 shows the performance of the MCI with only two agents (agents 1 and 3) being considered. A similar phenomenon can also be found in Table 3 where the more agents being considered leads to better performance. Interestingly, the opposite phenomenon is found for agent 2 where the APE is reduced from 5.33 (Table 1) to 4.83 meters (Table 2) using GNSS/VINS/OVD integration.

In short, based on the evaluated results in urban canyon 1, we conclude that:

(1) The proposed MCI framework can achieve significantly improved performance compared with the single-agent

TABLE 2. Positioning performances of MCI based on agents 1 and 2 using different methods in urban canyon 1.

Methods	Agent 1		Agent 2	
	MEAN	STD	MEAN	STD
GNSS/VINS/OVD	5.02 m	3.12 m	4.83 m	2.04 m
GNSS/VINS/OVD/IR	4.03 m	2.26 m	4.44 m	1.93 m

TABLE 3. Positioning performances of MCI based on agents 1 and 4 using different methods in urban canyon 1.

Methods	Agent 1		Agent 3	
	MEAN	STD	MEAN	STD
GNSS/VINS/OVD	4.67 m	3.02 m	5.11 m	2.65 m
GNSS/VINS/OVD/IR	4.41 m	2.53 m	5.01 m	2.74 m

positioning method, such as GNSS, VINS, and the GNSS/VINS integration.

(2) Compared with the overlapping area detection-based MCI, the inter-ranging measurements can effectively improve the performance of MCI. An APE of about 5 meters is obtained based on the GNSS/VINS/OVD/IR-based MCI.

(3) With the significantly increased amount of measurements from multiple agents, effectively identify the fault measurements is significant for improving the overall positioning accuracy of the agents.

B. EXPERIMENTAL EVALUATION IN URBAN CANYON 2

In this experiment, sensor measurements are provided by three agents. The scene of the evaluated urban canyon 2 can be seen in Fig. 7. The positioning performances of the listed five methods are shown in Table 4. In this experiment, data from 3 agents are integrated using the proposed MCI framework. Similar to the evaluation in urban canyon 1, both the single-agent positioning and multi-agents positioning solutions are compared. An APE of 14.69 meters is obtained based on GNSS standalone positioning for agent 1 with a standard deviation of 11.89 meters. We can see that the APE is larger than the one (12.49 meter of agent 1) in urban canyon 1. Similar performances are obtained for the other agents regarding the GNSS standalone positioning, as the GNSS measurements are collected from the same area. In short, we can see that the positioning error is larger than the one in urban canyon 1, due to the taller buildings in urban canyon 2, leading to severer GNSS NLOS and multipath effects.

The fourth row of Table 4 shows the relative positioning error of VINS. An RPE of 1.08 meters is achieved based on VINS for agent 1. Interestingly, the RPEs of all the agents are larger than the ones in urban canyon 1. The results show that urban canyon 2 is more challenging than the urban canyon 1, leading to larger positioning errors in VINS. Meanwhile, we can still find that the VINS can provide

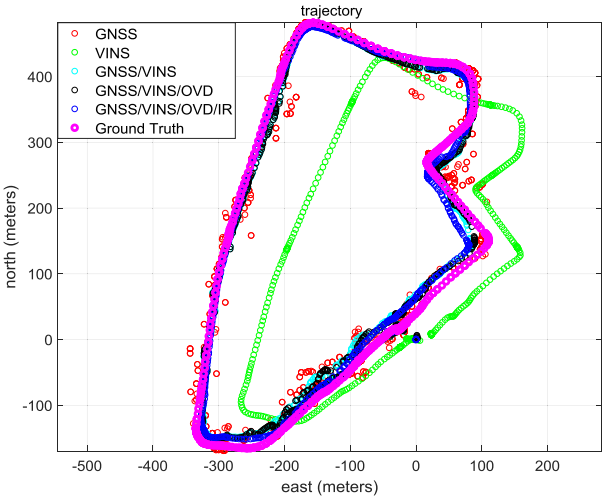


FIGURE 11. Illustration of the trajectories of the listed five methods for agent 1 during the evaluated dataset collected in urban canyon 2. The x-axis denotes the north direction during the test. The y-axis denotes the east direction. The red curve denotes the trajectory of GNSS standalone positioning. The green curve denotes the VINS solution. The cyan curve represents the trajectory of GNSS/VINS integration. The black and blue curves denote the trajectories based on proposed MCI frameworks, respectively. The magenta curve denotes the reference trajectory from SPAN-CPT.

more accurate positioning relatively in a short period, compared with the GNSS positioning. Similarly, the fifth row of Table 4 shows the APE of VINS, which denotes the error accumulation. An APE of 33.13 meters is obtained from VINS for agent 1 with an STD of 11.02 meters. Interestingly, we can see that the APE is even larger than the APE of GNSS standalone positioning (14.69 meters). Moreover, both the APEs for agent 3 and 4 reaches more than 17 meters.

Similar to the validation in urban canyon 1, we also show the performance of the integration of VINS and GNSS in the sixth row of Table 4. Interestingly, the accuracies are also slightly improved compared with both the VINS or GNSS standalone positioning for all the agents, which again show the complementariness of VINS and GNSS. An APE of 13.75 meters is obtained for the agent 1 with an STD of 6.69 meters. The APE is significantly smaller than both the GNSS (14.69 meters) and VINS (33.13 meters) standalone positioning. A similar phenomenon can also be seen by other agents. However, the remaining APE still reaches more than 10 meters for all the agents using GNSS/VINS integration.

With the help of the overlapping area detection constraint using the proposed MCI framework, the APE is decreased from 13.75 meters to 12.64 meters. Interestingly, the STD is slightly increased from 6.69 to 6.99 meters. Similar improvements can also be seen in the other agents. The improved accuracy shows that the proposed method which involves the OVD can help to improve the positioning performance of all the agents. Similarly, the main reason behind the improvement is that the OVD constraints connects all the states of the agents and corrects the corresponding potential errors by solving (12). However, we can see that the improvements

TABLE 4. Positioning performances of different agents using different methods in urban canyon 2.

Methods		Agent 1		Agent 2		Agent 3	
		MEAN	STD	MEAN	STD	MEAN	STD
Single Agent Positioning	GNSS	14.69 m	11.89 m	14.60 m	14.19 m	15.57 m	9.24 m
	VINS (RPE)	1.08 m	1.06 m	1.45 m	1.15 m	0.95 m	0.71 m
	VINS (APE)	33.13 m	11.02 m	29.17 m	13.72 m	17.01 m	13.59 m
	GNSS/VINS	13.75 m	6.69 m	13.26 m	5.71 m	10.34 m	4.14 m
Multi-Agents Positioning	GNSS/VINS/OVD	12.64 m	6.99 m	13.04 m	5.61 m	11.16 m	4.88 m
	GNSS/VINS/OVD/IR	11.87 m	7.36 m	9.77 m	6.48 m	7.04 m	4.67 m

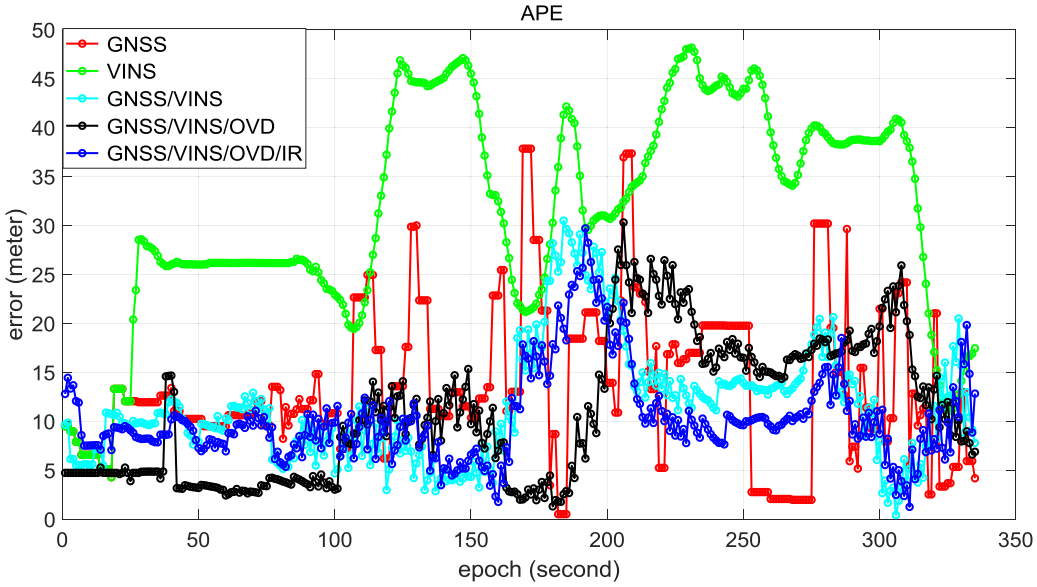


FIGURE 12. Illustration of the performance of the listed five methods for agent 1 during the evaluated dataset collected in urban canyon 2. The x-axis denotes the epoch during the test. The y-axis denotes the APE.

of the OVD constraint for all agents in urban canyon 2 are smaller than the one in urban canyon 1. One of the major reasons for this is the challenging scene which leads to more misdetection of OVD constraints. As we can see from Fig. 7, numerous dynamic objects lead to additional challenges to the applied OVD detection algorithm presented in Section IV-B. Moreover, the data is collected during the night which introduces additional challenges for OVD detection. To the best of the author's knowledge, this is the first paper that validates the OVD constraint in collaborative positioning using a dataset collected in the night period.

Effectively identifying the OVD in the night scene is still a challenging problem. Fortunately, with the help of the applied inter-ranging measurements, the APE is significantly improved for almost all the agents. The APE is decreased from 12.64 to 11.87 meters for agent 1. Larger improvements are achieved for agent 2 and 3. In short, the inter-ranging measurement can effectively improve the overall performance of all the agents even when the OVD constraints are limited, which shows the effectiveness of the proposed method. We are fully aware that the inter-ranging measurements in practice can be different than the ones that simulated in this paper. In the next step, we will conduct the real experiment to validate the effectiveness of the proposed MCI framework and it is interesting to see the difference between then.

To show the detail of the performance of a single agent using the listed five methods during the evaluation, we present the APEs of agent 1 in Fig. 12. Like Fig. 8, the red and green curves denote the APE of GNSS and VINS standalone positioning, respectively. The GNSS positioning fluctuates dramatically during the test, with the maximum APE reaching more than 35 meters. The APE of VINS is significantly smoother compared with the GNSS positioning. The cyan curve denotes the APE of GNSS/VINS integration with significantly smaller errors. However, the maximum error can still reach more than 30 meters. Interestingly, we can see that the GNSS/VINS integration is significantly smoother in urban canyon 1. One of the major reasons is that the GNSS positioning noise in urban canyon 1 is simulated using Gaussian distribution. However, the data of GNSS in urban canyon 2 is collected using the real GNSS receiver and the GNSS signal is severely affected by the NLOS effects. According to our findings in [52], the noise of GNSS tends to be a Gaussian mixture model-based distribution. However,

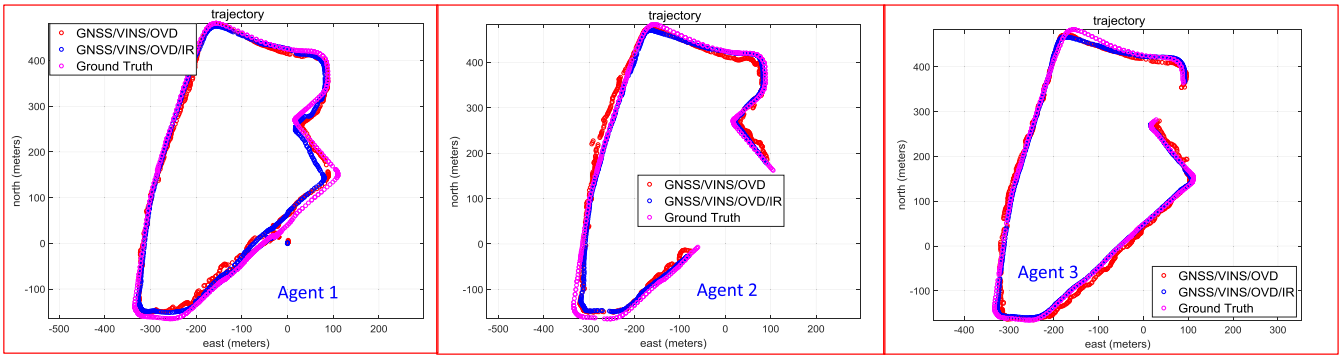


FIGURE 13. Illustration of the trajectories of four agents using the proposed MCI frameworks in urban canyon 1.

FIGURE 14. Experimental setup for future experimental validation which involves real inter-ranging measurements and message transmission with time delay.

the FGO relies heavily on the assumption of Gaussian noise of sensor measurement. Therefore, we believe that the violation of the Gaussian assumption is one of the major reasons leading to more fluctuant errors of GNSS/VINS integration in urban canyon 2.

The black and blue curves denote the APE of both the GNSS/VINS/OVD and the GNSS/VINS/OVD/IR integrations, respectively. The performance is improved almost throughout the evaluated dataset compared with the single agent-based positioning, which shows the effectiveness of the proposed MCI framework. Similar to Fig. 9, Fig. 13 shows the trajectories of the listed five methods for agent 1. We can see that the trajectory of the VINS (green curve) deviates significantly from the reference trajectory. The GNSS positioning fluctuates dramatically from the reference trajectory (magenta curve). The proposed MCI framework achieved smoother and more accurate trajectories concerning the reference trajectory, which again shows the effectiveness of the proposed MCI framework.

Interestingly, the APE of GNSS/VINS integration is even smaller than the proposed MCI framework at epoch 180. The main reason behind it is the misdetection of the OVD constraint. the misdetection of OVD can introduce a large positioning error to the whole factor graph shown in Fig. 2. As mentioned previously, the night dataset in urban canyon 2 introduces significantly larger challenges for OVD constraints detection.

Fig. 13 shows the trajectories of four agents using the proposed MCI frameworks (both the GNSS/VINS/OVD and GNSS/VINS/OVD/IR integrations). We can see that the agent 1 possess the longest trajectory with the overlapping areas with all the other three agents. In short, we can find that the best performance is obtained using the GNSS/VINS/OVD/IR integrations which again shows the effectiveness of the proposed method.

VI. DISCUSSION AND CONCLUSION

Achieving accurate positioning from the low-cost onboard sensors (GNSS/IMU/camera) from a single agent in the urban canyon is still a challenging problem to be solved. Instead of estimating the vehicular state simply based on onboard sensors from a single agent, the paper proposes a multi-agents collaborative integration (MCI) framework to improve the positioning accuracy, based on onboard sensors and leverage both the vision-based overlapping area detection and the inter-ranging measurements to boost the cross-connection between multi-agents. We validate the proposed MCI framework using challenging datasets collected in urban canyons of Hong Kong. The evaluated results show the effectiveness of the proposed MCI framework compared with the single agent-based positioning.

Based on the validation in urban canyon 2, we find that the performance of OVD detection is significantly challenged in the night scene. However, the inter-ranging measurements can effectively boost the cross-connection between agents, even when the OVD constraint is limited. In the future work, we will study to explore the advantage of massive data in MCI to detect the potential outlier measurements, for example, the misdetection of OVD, GNSS outliers.

Since the inter-ranging measurements are simulated based on the work in [21], and the time-delay of message transmission between the agent and centralized station is not considered in this paper, we will conduct the experiments which involve real inter-ranging measurements and message transmission with time delay. The experimental setup is shown in Fig. 14 which is provided by the Institute of Software Application Technology, Guangzhou.

AUTHOR CONTRIBUTIONS

Conceptualization, Weisong Wen, Guohao Zhang, and Li-Ta Hsu; methodology and software, Xiwei Bai and Weisong

Wen; paper review, and discussion for future experimental validation, Shendong Chen and Feng Yuan.

REFERENCES

[1] G. Dimitrakopoulos and P. Demestichas, "Intelligent transportation systems," *IEEE Veh. Technol. Mag.*, vol. 5, no. 1, pp. 77–84, Mar. 2010.

[2] P. Barber and N. Clarke, "Advanced collision warning systems," in *Proc. IEE Colloq. Ind. Automat. Control, Appl. Automot. Ind.*, London, U.K., 1998.

[3] D. Eckhoff, B. Halmos, and R. German, "Potentials and limitations of green light optimal speed advisory systems," in *Proc. IEEE Veh. Netw. Conf.*, Dec. 2013, pp. 103–110.

[4] K. Liu, S. H. Son, V. C. S. Lee, and K. Kapitanova, "A token-based admission control and request scheduling in lane reservation systems," in *Proc. 14th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2011, pp. 1489–1494.

[5] L.-T. Hsu, "Analysis and modeling GPS NLOS effect in highly urbanized area," *GPS Solutions*, vol. 22, no. 1, p. 7, Jan. 2018.

[6] L.-T. Hsu, Y. Gu, Y. Huang, and S. Kamijo, "Urban pedestrian navigation using smartphone-based dead reckoning and 3-D map-aided GNSS," *IEEE Sensors J.*, vol. 16, no. 5, pp. 1281–1293, Mar. 2016.

[7] L.-T. Hsu, Y. Gu, and S. Kamijo, "3D building model-based pedestrian positioning method using GPS/GLONASS/QZSS and its reliability calculation," *GPS Solutions*, vol. 20, no. 3, pp. 413–428, Jul. 2016.

[8] W. Wen, G. Zhang, and L.-T. Hsu, "GNSS NLOS exclusion based on dynamic object detection using LiDAR point cloud," *IEEE Trans. Intell. Transp. Syst.*, early access, Dec. 31, 2019, doi: 10.1109/TITS.2019.2961128.

[9] W. Wen, G. Zhang, and L. Hsu, "Correcting NLOS by 3D LiDAR and building height to improve GNSS single point positioning," *Navigation*, vol. 66, no. 4, pp. 705–718, Dec. 2019.

[10] L.-T. Hsu, Y. Gu, and S. Kamijo, "NLOS correction/exclusion for GNSS measurement using RAIM and city building models," *Sensors*, vol. 15, no. 7, pp. 17329–17349, Jul. 2015.

[11] W. Wen, X. Bai, Y. C. Kan, and L.-T. Hsu, "Tightly coupled GNSS/INS integration via factor graph and aided by fish-eye camera," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10651–10662, Nov. 2019.

[12] P. D. Groves, *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*. Norwood, MA, USA: Artech House, 2013.

[13] K.-W. Chiang, G.-J. Tsai, H.-J. Chu, and N. El-Sheimy, "Performance enhancement of INS/GNSS/Refreshed-SLAM integration for acceptable lane-level navigation accuracy," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 2463–2476, Mar. 2020.

[14] H. Xiong, R. Bian, Y. Li, Z. Du, and Z. Mai, "Fault-tolerant GNSS/SINS/DVL/CNS integrated navigation and positioning mechanism based on adaptive information sharing factors," *IEEE Syst. J.*, early access, Apr. 7, 2020, doi: 10.1109/JSYST.2020.2981366.

[15] T. Qin, P. Li, and S. Shen, "VINS-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, Aug. 2018.

[16] X. Bai, W. Wen, and L.-T. Hsu, "Robust visual-inertial integrated navigation system aided by online sensor model adaption for autonomous ground vehicles in urban areas," *Remote Sens.*, vol. 12, no. 10, p. 1686, May 2020.

[17] X. Bai, W. Wen, and L.-T. Hsu, "Performance analysis of visual/inertial integrated positioning in typical urban scenarios of Hong Kong," in *Proc. Asian–Pacific Conf. Aerosp. Technol. Sci.*, 2019, pp. 1–9.

[18] T. Qin, J. Pan, S. Cao, and S. Shen, "A general optimization-based framework for local odometry estimation with multiple sensors," 2019, *arXiv:1901.03638*. [Online]. Available: http://arxiv.org/abs/1901.03638

[19] M. Kais, L. Bouraoui, S. Morin, A. Porterie, and M. Paren, "A collaborative perception framework for intelligent transportation system applications," in *Proc. 12th World Congr. Intell. Transp. SystemsITS AmericaITS JapanERTICO*, 2005, p. 12.

[20] H. Ko, B. Kim, and S.-H. Kong, "GNSS multipath-resistant cooperative navigation in urban vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, pp. 5450–5463, Dec. 2015.

[21] S. Tanwar and G. X. Gao, "Multi-epoch multi-agent collaborative localization using grid-based 3DMA GNSS and inter-agent ranging," in *Proc. 32nd Int. Tech. Meeting Satell. Division Inst. Navigat. (ION GNSS+)*, Miami, FL, USA, Oct. 2019, pp. 1963–1975.

[22] N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, and R. J. O'Dea, "Relative location estimation in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2137–2148, Aug. 2003.

[23] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 54–69, Jul. 2005.

[24] N. Drawil and O. Basir, "Vehicular collaborative technique for location estimate correction," in *Proc. IEEE 68th Veh. Technol. Conf.*, Sep. 2008, pp. 1–5.

[25] Y. T. Chan, W. Y. Tsui, H. C. So, and P.-C. Ching, "Time-of-Arrival based localization under NLOS conditions," *IEEE Trans. Veh. Technol.*, vol. 55, no. 1, pp. 17–24, Jan. 2006.

[26] F. Gustafsson and F. Gunnarsson, "Positioning using time-difference of arrival measurements," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 6, Apr. 2003, p. VI-553.

[27] C. Feng, W. S. A. Au, S. Valaee, and Z. Tan, "Received-signal-strength-based indoor positioning using compressive sensing," *IEEE Trans. Mobile Comput.*, vol. 11, no. 12, pp. 1983–1993, Dec. 2012.

[28] K. Liu, H. B. Lim, E. Frazzoli, H. Ji, and V. C. S. Lee, "Improving positioning accuracy using GPS pseudorange measurements for cooperative vehicular localization," *IEEE Trans. Veh. Technol.*, vol. 63, no. 6, pp. 2544–2556, Jul. 2014.

[29] F. Dellaert and M. Kaess, "Factor graphs for robot perception," *Found. Trends Robot.*, vol. 6, nos. 1–2, pp. 1–139, 2017.

[30] N. Alam, A. Tabatabaei Balaei, and A. G. Dempster, "Relative positioning enhancement in VANETs: A tight integration approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 47–55, Mar. 2013.

[31] G. Zhang, W. Wen, and L.-T. Hsu, "Rectification of GNSS-based collaborative positioning using 3D building models in urban areas," *GPS Solutions*, vol. 23, no. 3, p. 83, Jul. 2019.

[32] G. Zhang, H.-F. Ng, W. Wen, and L.-T. Hsu, "3D mapping database aided GNSS based collaborative positioning using factor graph optimization," *IEEE Trans. Intell. Transp. Syst.*, early access, Apr. 30, 2020, doi: 10.1109/TITS.2020.2988531.

[33] W. Wen, Y. Zhou, G. Zhang, S. Fahandezh-Saadi, X. Bai, W. Zhan, M. Tomizuka, and L.-T. Hsu, "UrbanLoco: A full sensor suite dataset for mapping and localization in urban scenes," 2019, *arXiv:1912.09513*. [Online]. Available: http://arxiv.org/abs/1912.09513

[34] H. Xu, L. Wang, Y. Zhang, K. Qiu, and S. Shen, "Decentralized visual-inertial-UWB fusion for relative state estimation of aerial swarm," 2020, *arXiv:2003.05138*. [Online]. Available: http://arxiv.org/abs/2003.05138

[35] J. Huai, G. Jóźków, C. Toth, and D. A. Grejner-Brzezinska, "Collaborative monocular SLAM with crowdsourced data," *Navigation*, vol. 65, no. 4, pp. 501–515, Dec. 2018.

[36] M. Karrer and M. Chli, "Towards globally consistent visual-inertial collaborative SLAM," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 1–8.

[37] D. Zou and P. Tan, "CoSLAM: Collaborative visual SLAM in dynamic environments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 354–366, Feb. 2013.

[38] C. Forster, S. Lynen, L. Kneip, and D. Scaramuzza, "Collaborative monocular SLAM with multiple micro aerial vehicles," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 3962–3970.

[39] T. Schneider, M. Dymczyk, M. Fehr, K. Egger, S. Lynen, I. Gilitschenski, and R. Siegwart, "Maplab: An open framework for research in visual-inertial mapping and localization," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 1418–1425, Jul. 2018.

[40] R. Mur-Artal and J. D. Tardos, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017.

[41] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2007, pp. 3565–3572.

[42] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *Int. J. Robot. Res.*, vol. 34, no. 3, pp. 314–334, 2015.

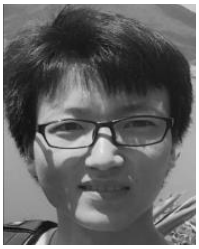
[43] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* Piscataway, NJ, USA: IEEE Press, Jun. 1994, pp. 593–600.

[44] T. Senst, V. Eiselein, and T. Sikora, "II-LK—A real-time implementation for sparse optical flow," in *Proc. Int. Conf. Image Anal. Recognit.* Springer, 2010, pp. 240–249.

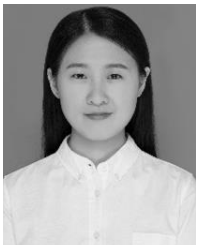
[45] S. Agarwal *et al. Ceres Solver*. [Online]. Available: http://ceres-solver.org

[46] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate $O(n)$ solution to the pnp problem," *Int. J. Comput. Vis.*, vol. 81, no. 2, p. 155, 2009.

[47] D. Galvez-López and J. D. Tardos, "Bags of binary words for fast place recognition in image sequences," *IEEE Trans. Robot.*, vol. 28, no. 5, pp. 1188–1197, Oct. 2012.
[48] P. Sun *et al.*, "Scalability in perception for autonomous driving: Waymo open dataset," 2019, *arXiv:1912.04838*. [Online]. Available: http://arxiv.org/abs/1912.04838
[49] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: An open-source robot operating system" in *Proc. ICRA Workshop Open Source Softw.*, Kobe, Japan, 2009, vol. 3, no. 3.2, p. 5.
[50] M Grupp. (2017). *Evo: Python Package for the Evaluation of Odometry and Slam*. [Online]. Available: https://github.com/MichaelGrupp/evo
[51] N. Sunderhauf and P. Protzel, "Switchable constraints for robust pose graph SLAM," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 1879–1884.
[52] W. Wen, T. Pfeifer, X. Bai, and L.-T. Hsu, "It is time for factor graph optimization for GNSS/INS integration: Comparison between FGO and EKF," 2020, *arXiv:2004.10572*. [Online]. Available: http://arxiv.org/abs/2004.10572



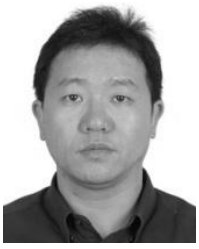
WEISONG WEN was born in Ganzhou, Jiangxi, China. He is currently pursuing the Ph.D. degree in mechanical engineering with The Hong Kong Polytechnic University. He was a Visiting Student Researcher with the University of California at Berkeley (UCB), in 2018. His research interests include multi-sensor integrated localization for autonomous vehicles, simultaneous localization and mapping, and global navigation satellite systems positioning in urban canyons.



XIWEI BAI is currently a Research Assistant with the Interdisciplinary Division of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University. Her research interests include global navigation satellite systems positioning aided by computer vision, object detection, and recognition using the deep neural networks.

GUOHAO ZHANG received the bachelor's degree in mechanical engineering and automation from the University of Science and Technology, Beijing, China, in 2015, and the master's degree in mechanical engineering from The Hong Kong Polytechnic University, where he is currently pursuing the Ph.D. degree. His research interests including global navigation satellite system urban localization, collaborative positioning, and multisensory integrated navigation.

SHENGDONG CHEN is currently the Deputy Director of the Intelligent Transportation Laboratory, Institute of Software Application Technology, Guangzhou Chinese Academy of Sciences. His research interests include cloud computing, data mining, pattern recognition, deep learning, the Internet of Things, autopilot vehicle applications, cooperative vehicle infrastructure, and intelligent transportation.



FENG YUAN is currently pursuing the Ph.D. degree with the Institute of Software Application Technology, Guangzhou Chinese Academy of Sciences. He is also the Executive Deputy Director of the Institute of Software Application Technology, Guangzhou Chinese Academy of Sciences. He is also the Deputy Director of the Software Development Research Department, Institute of Software Development, Chinese Academy of Sciences. He is also an Expert with the Intelligent City Industry Alliance Ministry of Industry and Informatization and the Ministry of Intelligent City Housing and Construction. His research interests include the Internet of Things, software engineering, intelligent city, application of intelligent common pole, and city level big-data platform of the Internet of Things.

LI-TA HSU (Member, IEEE) received the B.S. and Ph.D. degrees in aeronautics and astronautics from National Cheng Kung University, Taiwan, in 2007 and 2013, respectively. He was a Postdoctoral Researcher with the Institute of Industrial Science, University of Tokyo, Japan. He was a Visiting Scholar with University College London, U.K., in 2012. He is currently an Assistant Professor with the Interdisciplinary Division of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University. His research interests include global navigation satellite system positioning in challenging environments and localization for pedestrian, autonomous driving vehicle, and unmanned aerial vehicle.