Earthq Eng & Eng Vib (2023) 22: 371-386 **DOI**: https://doi.org/10.1007/s11803-023-2175-y

Graph-based structural joint pose estimation in non-line-of-sight conditions

Alexander Thoms[1†], Zaid Al-Sabbag[2‡] and Sriram Narasimhan[3§]

1. *Department of Mechanical and Aerospace Engineering, University of California – Los Angeles, USA*
2. *Department of Civil and Environmental Engineering, University of Waterloo, Canada*
3. *Department of Civil and Environmental Engineering, University of California – Los Angeles, USA*

Abstract: In post-earthquake surveys, it is difficult (and often infeasible) to observe and quantify displacements beyond line-of-sight (LOS), given seismic force-resisting and gravity systems exist completely or partially within a building′s enclosure. To overcome this limitation, we develop a novel framework that generalizes graph-based state estimation towards structural joint localization via engineered landmarks. These landmarks provide an indirect means to estimate residual displacements where direct LOS is unavailable. Within our framework, engineered landmarks define topologies of uniquely identifiable landmarks that are either visible or non-visible to a robot performing simultaneous localization and mapping (SLAM). Within the SLAM approach, factors encoding robot odometry and robot-to-visible landmark measurements are formulated for the cases of wireless sensing and fiducial object detection and tracking. Visible landmarks are rigidly attached to non-visible landmark subsets for each engineered landmark, where the complete set of non-visible landmarks form globally rigid and localizable connectivity graphs via range-based factors. Complimentary subsets of non-visible landmarks are embedded within the base structure and uniquely define joint pose via geometric factors. All factors are unified within a common graph to solve for the maximum *a posteriori* estimate of robot, landmark, and joint states via nonlinear least squares optimization. To demonstrate the applicability of our approach, we apply the Monte Carlo method over a parameterization of system noise to calculate residual joint pose error distributions, maximum average inter-story drift ratios, and related summary statistics for a 19-story nonlinear structural model. By performing nonlinear time history analyses over sets of service-level and maximum considered earthquakes, our parametric study gives insight into our method′s application towards post-earthquake building evaluation in non-LOS conditions.

Keywords: state estimation; engineered landmarks; post-earthquake; residual displacement

1 Introduction

Post-earthquake, damage assessment is a necessary step towards the safe demolition, repair, and/or reoccupation of critical infrastructure. Standard methods of assessment, such as those described in ATC-20 (Applied Technology Council, 2005), centre on inspectors evaluating exterior, visual signs of structural and non-structural component damage to infer the extent of remaining structural capacity. With inference based on human judgment and subjective ratings of component damage, building safety is consequently limited to categorical designations, such as: safe to reoccupy, limited entry, and unsafe to reoccupy. Research has looked towards augmenting these practices to explicitly measure component damage and to relate these damage states to building safety. Methods proposed for measuring component damage in situ are numerous and generally centre on structural health monitoring (SHM) and remote/robotic inspection. We acknowledge predictive modelling (Zhang *et al.*, 2018; Zhang and Burton, 2019; Zhang *et al.*, 2019) too plays a significant role in post-earthquake damage assessment, though in this paper we limit our review to on-site methods.

Among other damage measures, traditional SHM has looked towards measuring inter-story drift as a global engineering demand parameter (EDP) given its wide adoption in seismic design codes and standards (ASCE/SEI 43-05, 2005; ASCE/SEI 7-16, 2016; Eurocode 8, 2004; NZS 1170.5, 2004). More specifically, seismic design codes and standards categorize damage and performance limit states proportional to the inter-story drift ratio (IDR), which is defined as the relative translational displacement between two consecutive floors divided by the story height (Skolnik and Wallace, 2010). To objectively estimate IDR for post-earthquake damage assessment, early works focused on massively

Correspondence to: Sriram Narasimhan, 580 Portola Plaza, 5731 Boelter Hall, Los Angeles, CA 90095-159, USA
E-mail: snarasim@ucla.edu
[†]PhD Student; [‡]PhD Candidate; [§]Professor
Received October 17, 2022; **Accepted** February 11, 2023

instrumenting buildings with accelerometers and wire gauges, such as linear variable differential transmitters (LVDT) (Celebi, 2005). With this instrumentation, acceleration measurements on consecutive floors are double integrated to recover IDR while the lengthening/shortening of wires (spanning diagonally across framing) is used to directly estimate IDR assuming rigid center-line motions. The simplicity of these methods comes at a cost of accuracy and reliability, with inelastic deformations under-estimated via high-pass filtering of accelerations and long span frames, rotations, wire creep/sag, and partition walls reducing the effectiveness of wire gauge readings (Skolnik and Wallace, 2010).

Modern SHM methods have looked towards the application of computer vision (Park *et al.*, 2015; Hsu and Kuo, 2020; Yang *et al.*, 2020b) and optical laser sensors (McCallen *et al.*, 2017; Petrone *et al.*, 2018) to directly measure transient and residual IDR for accurate and repeatable observations. Park *et al.* (2015) employ a motion capture system to estimate 3D structural displacements (via the tracking of markers defining structural joint pose) and demonstrate their method effective under torsional and lateral forcing. The author′s method, however, is limited to small scale testing and neglects non-line-of-sight (non-LOS) conditions, rendering this approach practically infeasible in the assessment of internal damage. In scaling computer vision to real-world application, Hsu and Kuo (2020) presents a stand-alone smart camera system that can be deployed within a building′s seismic force-resisting system (SFRS) to directly estimate IDR. The system compensates for camera rotations during earthquake excitation as well as residual IDR via fusion of accelerometer and camera-based IDR measurements. Yang *et al.* (2020b) develops a similar stand-alone system, though forgoes the use of accelerometers and mounts the camera outside of the building′s SFRS to track discernible features within ceilings while additionally measuring torsional responses. Using optical sensors, McCallen *et al.* (2017) and Petrone *et al.* (2018) propose directly measuring IDR across consecutive floors; however, assumptions on SFRS configuration and in-plane rotations limit these systems to specific structural configurations. Generalizing IDR estimation to arbitrary structural configurations can be approached via joint translation measurements, where the additional measure of joint rotations potentiates the use of local EDPs (such as frame beam rotation, frame column rotation, etc.) for a more holistic approach to building safety (Yazgan and Dazio, 2012; Zhang *et al.*, 2019).

Driven by the above need and technical challenge in estimating six degree-of-freedom (DoF) structural joint displacements, we develop a graph-based state estimation framework that allows for indirect measure of residual joint displacements for post-earthquake damage assessment. Central to our approach is simultaneous localization and mapping (SLAM), which addresses the problem of a robot acquiring a spatial map of an environment while simultaneously localizing itself relative to the generated map (Stachniss *et al.*, 2016). To date, remote/robotic solutions for post-earthquake building evaluation have focused on the collection of visual data to inform decision making, with SLAM (Mao *et al.*, 2018; Recchiuto and Sgorbissa, 2018), structure-from-motion (Torok *et al.*, 2012, 2014), and digital-twin techniques (Levine and Spencer, 2022; Wang *et al.*, 2022) providing a means for quantitative assessment. Similar to manual inspections, damage is inferred from visual information only with solutions yet to integrate methods of estimating in situ structural deformations. To address this research gap, our framework applies graph-based state estimation to structural joint localization via engineered landmarks. Engineered landmarks define topologies of uniquely identifiable landmarks that are either visible or non-visible to the robot performing SLAM. Within the SLAM approach, factors encoding robot odometry and robot-to-visible landmark measurements are formulated for the cases of wireless sensing and fiducial marker detection and tracking (herein referred to as fiducial objects). Active wireless sensors (herein referred to as anchors) act within a wireless sensor network (WSN) to sense adjacent nodes via time of flight (ToF) range measurements with known correspondence (Yick *et al.*, 2008) while fiducial objects provide an easily recognizable feature with embedded fault detection for robust pose estimation (Kalaitzakis *et al.*, 2021). These visible landmarks are rigidly attached to non-visible landmarks subsets for each engineered landmark, where the complete set of non-visible landmarks form globally rigid and localizable connectivity graphs (Priyantha *et al.*, 2005) via range-based factors that are generalized to any range-based sensor modality. Complimentary subsets of non-visible landmarks are embedded within the base structure and uniquely define joint pose via geometric factors. Joint pose is recovered by solving the maximum *a posteriori* (MAP) estimate of robot, landmark, and joint states in a nonlinear least squares optimizer. In the advent of the Internet of Things (Khelifi *et al.*, 2019), we anticipate centimeter-level accurate wireless sensing technology (via ultra-wideband (UWB), mm-Wave, etc.) that is both affordable and low-power, allowing wireless sensors to be economically deployed at scale. Furthermore, the use of linear displacement sensors and compact LVDTs potentiates sub-millimeter accurate range measurements from visible to non-visible landmarks. Figure 1 illustrates our proposed framework, which is applicable to arbitrary structural configurations.

The remaining sections of this paper are as follows: Section 2 discusses related work; Section 3 presents our problem formulation with a primer on factor graphs and related nonlinear machinery; Section 4 details experiments conducted to validate our framework; Section 5 concludes with our outlook on future work.

2 Related works

To our knowledge, there is no work that addresses structural joint pose estimation in non-LOS conditions. Given this, we focus our review on methods enabling our approach, which include sensor fusion strategies enabling robot-to-visible landmark measurements and rigid body localization methods enabling joint pose estimation.

2.1 Sensor fusion strategies

Wireless sensor fusion and fiducial object detection and tracking methods have demonstrated improvements to SLAM system accuracy and robustness, with both modalities affording unique landmark localization with known correspondence. For wireless sensing, UWB has been studied extensively due to its robustness against multipath and non-LOS effects while maintaining relatively high ToF accuracy among alternative wireless technologies (Aiello and Rogerson, 2003). While we emphasize UWB fusion in our review, sensor fusion strategies generalize to any ToF wireless sensing modality. Literature on fiducial object detection and tracking is extensive as well, and we briefly review a key study quantifying fiducial measurement model accuracy as the inclusion of fiducial object detection and tracking in visual SLAM is handled similarly to visual keypoint detection and tracking (Munoz-Salinas and Medine-Carnicer, 2020). Note that we limit our review on wireless sensor fusion to graph-based methods, which tend to perform better than filter-based methods (Dellaert and Kaess, 2017).

2.1.1 Wireless sensor fusion

Wang *et al.* (2017) present one of the first studies which fuses UWB with graph-based SLAM. The authors employ a loosely-coupled approach where UWB localization, visual-inertial odometry (VIO), and map optimization operate in parallel threads. Locally consistent odometry from VIO and drift-free global position estimates from UWB localization are added to a pose graph in the map optimization thread, where bundle adjustment is performed over a sliding window. As part of the optimization, cost functions formalizing range and smoothing errors are added, with range errors assumed zero mean Gaussian without bias. The authors observe their hand-crafted smoothing error, constraining consecutive keyframes via a motion model, to improve trajectory smoothness, compensating for noise introduced by the UWB localization thread.

Lutz *et al.* (2019) follow a similar loosely-coupled approach, extending their graph to optimize over anchor landmark position while jointly calibrating a UWB sensor measurement error model for all anchor-anchor pairs. The author′s error model accounts for antenna directivity, antenna delay bias, and biases introduced by non-deterministic signal propagation (i.e., non-LOS conditions), with the latter observed to introduce the majority of outlier measurements in testing. Upon calibration, error models are simplified to Gaussian distributions with mean values representing inherent biases and variances approximating ranging uncertainty after outlier removal. Biases from calibration are then added to UWB range factors, with residual errors

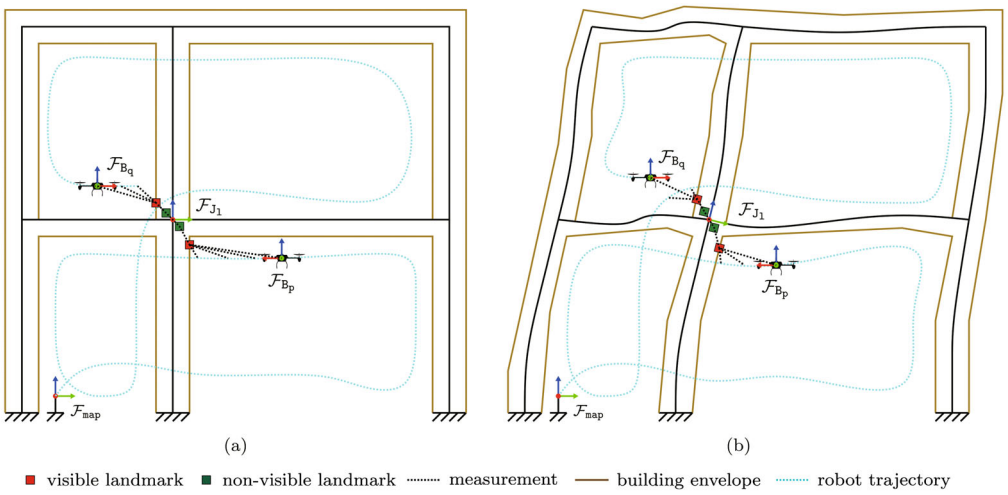

Fig. 1 Conceptual diagram of our six degree-of-freedom structural joint pose estimation framework in non-line-of-sight conditions. To estimate residual joint displacements, robotic inspections (a) pre- and (b) post-earthquake are initialized from frame $\mathcal{F}_{map}$, which is defined by a landmark (ex. fiducial object) and is assumed globally consistent pre- and post-event. During a single simultaneous localization and mapping session, the robot navigates through each floor to measure all visible landmarks (red), enabling the MAP estimate of structural joint pose given non-visible landmarks (green) constrain joint and visible landmark states. For the particular scenario illustrated, the robot measures visible landmarks (from keyframes p and q within its body frame B) associated with joint J_t. Additional measurements from the robot to visible landmarks are made over the course of the robot′s trajectory to better determine engineered landmark location

weighted by a Cauchy loss function.

Following trends in tightly-coupled formulations, Nguyen *et al.* (2020a) propose a monocular visual odometry system aided by a single UWB anchor. In this work, visual odometry factors and UWB ranging factors (similar to Wang *et al.* (2017)) are optimized jointly in a pose graph. Key contributions include 1) an initialization procedure leveraging UWB ranging measurements to recover metric scale and anchor position and 2) a monitoring process for refining these estimates over the course of data collection. The authors provide a framework to verify the existence of a solution to their initialization procedure, noting effective initialization when 1) an initial estimate of anchor position with respect to the first camera frame is provided, 2) there is LOS from anchor to robot, and 3) the robot moves about multiple axes. The author′s monitoring system augments every keyframe by the nearest range measurement and estimated anchor position in a sliding window, minimizing both visual reprojection errors and UWB ranging errors via bundle adjustment. Augmentation occurs when the error on the current anchor position exceeds a prescribed threshold, reinitializing the system. Building on their own work in (Nguyen *et al.*, 2020b), the authors extend their contributions to the handling of multiple anchors in degenerate configurations and the relaxation of good initial anchor position estimates via prior range measurements between anchor-world and anchor-anchor pairs.

Leading from (Nguyen *et al.*, 2020a, 2020b), tightly-coupled formulations focusing on visual-inertial-ranging (Cao and Beltrame, 2021; Nguyen *et al.*, 2021a), lidar-inertial-ranging (Nguyen *et al.*, 2021b), and lidar-visual-inertial-ranging (Nguyen *et al.*, 2021c) improve the state-of-the-art both in accuracy and robustness, with additional contributions for wireless fusion proposed. In (Nguyen *et al.*, 2021b), the authors leverage preintegrated inertial measurements to estimate the robot′s movement between consecutive keyframes. In doing so, UWB range factors associate each range measurement with a robot pose for a shared time point, compared to past formulations associating range measurements to keyframes within a pre-defined time threshold. This retains the full set of UWB measurements, improving system accuracy. In Nguyen *et al.* (2021c), the authors build their system around four UWB anchors (defining a world coordinate system) from which the 6-DoF robot pose is estimated. This departs from earlier works, where UWB range factors only constrain the robot′s position. When comparing their lidar-visual-inertial-ranging framework to visual-inertial-ranging and lidar-inertial-ranging configurations, absolute trajectory error is reduced across all scenarios on the NTU VIRAL data set (Nguyen *et al.*, 2021d).

2.1.2 Fiducial measurement modelling

Within visual SLAM systems, AprilTag (Olson, 2011) and Aruco (Romero-Ramirez *et al.*, 2018) fiducial objects can be readily detected in gray-scale imaging systems and tracked similarly to other visual key points. Towards improving the localization accuracy of these fiducial objects, Kallwies *et al.* (2020) present visual-processing techniques that 1) filter out inaccurate detections resulting from partial board occlusion and 2) refine detected edges and corners. With techniques jointly applied, median localization errors are reduced by a factor of ten to 0.017 px over existing implementations. The authors further quantify detection rate and RMSE pixel error as a function of incidence angle (between the imaging system and fiducial object) and distance from which we establish a simplified robot-to-fiducial object measurement model in Section 4.1.

2.2 Rigid body localization

A method for rigid body localization using a WSN is proposed by Chepuri *et al.* (2014). In this work, the authors formulate joint pose estimation via anchor-tag range measurements and tag-joint geometric relations in a nonlinear least-squares optimizer. Here, tags refer to passive wireless sensors that are sensed by anchors with known correspondence, where tag topology relative to the joint's coordinate frame potentiates joint localization. Uncertainty in tag position (relative to the joint) is handled by constrained total-least-squares estimators, with our framework accounting for uncertainty via a custom factor (constraining pose to position) derived by Wisth *et al.* (2023). Note that in our work, we consider all WSN nodes integrated with engineered landmarks to be anchors.

3 Problem formulation

Our structural joint pose estimation framework unifies SLAM innovations discussed in Section 2.1 with the rigid body localization approach discussed in Section 2.2 within a common graph-based solution. In doing so, we leverage the interrelations between all connected states (i.e., robot, landmark, joint) to provide an estimate that maximally agrees with all sensor information. To our knowledge, unifying these approaches has yet to be presented in literature and serves as the key novelty in this work. Though common knowledge in the robotics community, we briefly discuss factor graphs and related machinery before proceeding to mathematical formalisms. Notation follows similarly to (Wisth *et al.*, 2023; Nguyen *et al.*, 2021c).

3.1 Preliminaries

3.1.1 Factor graphs

A factor graph (see Fig. 2) is a bipartite graph $F=\left(V=\{\mathcal{U},\mathcal{V}\},E\coloneqq\mathcal{E}\right)$ that encodes measurements $z\in Z$ (either relative z_{ij} or absolute z_i) as factors $\phi_i\in\mathcal{U}$ over a set of variables $\boldsymbol{x}_j\in\mathcal{V}$ describing a particular

system′s state space (Dellaert and Kaess, 2017). Factor nodes are connected to variables nodes via undirected edges $\epsilon_{ij} \in \mathcal{E}$, which encode functional dependencies between factors and their connected variables such that:

$$\varepsilon_{ij} \to \left\{\phi_i, \boldsymbol{x}_j \in \mathcal{N}\left(\phi_i\right)\right\} \tag{1}$$

where $\mathcal{N}\left(\phi_i\right)$ defines the set of variables connected to factor ϕ_i via ϵ_{ij}. The product over all factors yields the global function:

$$\phi(\mathcal{X}) = \prod_i \phi_i\left(\mathcal{X}_i\right) \tag{2}$$

where $\phi(\mathcal{X})$ specifies the joint density over $\mathcal{X}$ as a product of factors $\phi_i\left(\mathcal{X}_i\right) \propto p\left(\mathcal{X}_i | \, Z_i\right)$. In SLAM, we seek state (i.e., variable) values that maximally agree with the information present in the uncertain measurements (Dellaert and Kaess, 2017) such that:

$$\mathcal{X}^\star = \arg\max_{\mathcal{X}} \phi(\mathcal{X}) \tag{3}$$

where $\mathcal{X}^\star$ is the MAP of $\mathcal{X}$. Assuming measurements are conditionally independent and corrupted by white Gaussian noise, Eq. (3) can be formulated as a nonlinear least squares minimization problem:

$$\mathcal{X}^\star = \arg\min_{\mathcal{X}} \sum_i \| \mathbf{r}_i \|^2_{\boldsymbol{\Sigma}_i} \tag{4}$$

where $\mathbf{r}_i = h_i\left(\mathcal{X}_i\right) - z_i$ is the residual error taken as the difference between the measurement function $h_i\left(\mathcal{X}_i\right)$ and measurement z_i, and $\| (\cdot) \|^2_{\boldsymbol{\Sigma}_i} = (\cdot)^\top \boldsymbol{\Sigma}_i^{-1} (\cdot)$ is the squared Mahalanobis distance with covariance matrix $\boldsymbol{\Sigma}_i$. In this formulation, $h_i\left(\mathcal{X}_i\right)$ and z_i must belong to vector spaces to ensure standard Gauss–Newton nonlinear optimization schemes can be applied to solve for Eq. (4).

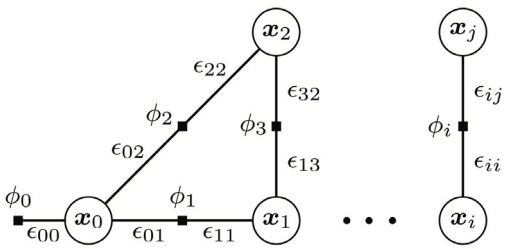

Fig. 2 A simple factor graph represented in its general form: $F = \left(V = \{\mathcal{U}, \mathcal{V}\}, E \coloneqq \mathcal{E}\right)$ (adapted from Dellaert and Kaess (2017)). Factors $\phi_i \in \mathcal{U}$ encode measurements over a set of variables $x_j \in \mathcal{V}$ describing a particular system′s state space, while edges $\epsilon_{ij} \in \mathcal{E}$ encode functional dependencies between factors and their connected variables

3.1.2 Machinery

Given $\mathcal{X}$ (as for SLAM and other robotics inference problems) contains rotations $\mathbf{R} \in \mathrm{SO}(3)$ that exist within a manifold $\mathcal{M}$, we reparameterize Eq. (3) such that:

$$\mathcal{X}^\star = \arg\max_{\delta\mathcal{X} \in \mathbb{R}^n} \phi\left(\mathcal{R}_{\mathcal{X}}\left(\delta\mathcal{X}\right)\right) \tag{5}$$

where $\mathcal{R}_{\mathcal{X}}(\cdot): \mathbb{R}^n \to \mathcal{M}$ is called a local reparameterization at $\mathcal{X}$, or retraction function (Nguyen *et al.*, 2021c). After the reparameterization, for each iteration in the Gauss–Newton method, the optimal gradient $\delta\mathcal{X}^\star$ is calculated and then the solution $\delta\mathcal{X}^\star$ from the tangent space is "retracted" back to the manifold using the operation $\mathcal{X} \leftarrow \mathcal{R}_{\mathcal{X}}\left(\delta\mathcal{X}^\star\right)$ (Forster *et al.*, 2016). We refer the interested reader to (Forster *et al.*, 2016; Nguyen *et al.*, 2021c) for details concerning retraction functions, Jacobians for functions acting on manifolds, and Riemannian geometry.

3.2 Proposed framework

Our framework′s state space is presented first, with landmark topology and MAP estimation formalisms following. A factor graph representation of our framework is illustrated in Fig. 3.

3.2.1 State space

Let the complete history of observed states $\mathcal{X}_k$ for all observable points up to time t_k be generalized as:

$$\mathcal{X}_k \triangleq \left\{\boldsymbol{x}_i, \boldsymbol{f}_\ell \triangleq \left\{\boldsymbol{v}_\ell, \boldsymbol{h}_\ell, \boldsymbol{J}_\ell\right\}\right\}_{i \in \mathrm{K}_k, \ell \in \mathrm{L}_k} \tag{6}$$

where $\boldsymbol{x}_i$ is the robot′s state at time t_i, with i belonging to the set of keyframes K_k; $\boldsymbol{f}_\ell$ is landmark state for a unique identifier ℓ which belongs to the set of landmarks $\mathrm{L}_k \triangleq \left\{\mathrm{V}_k, \mathrm{H}_k, \mathrm{J}_k\right\}$; V_k is the set of landmarks visible to the robot (i.e., LOS conditions exist from $\boldsymbol{x}_i \in \mathcal{X}_k$ to $\boldsymbol{v}_\ell \in \mathrm{V}_k$ for $\left|\{i\}\right| \geq 1$), H_k represents the set of landmarks non-visible to the robot (i.e., non-LOS conditions exist from $\boldsymbol{x}_i \in \mathcal{X}_k$ to $\boldsymbol{h}_\ell \in \mathrm{H}_k$ for all i), and $\boldsymbol{J}_\ell \in \mathrm{SE}(3) \in \mathrm{J}_k$ is joint state. For our study, we consider anchors $\boldsymbol{a}_\ell \in \mathbb{R}^3 \in \mathrm{A}_k$ and fiducial objects $\boldsymbol{F}_\ell \in \mathrm{SE}(3) \in \mathrm{F}_k$ as visible landmarks:

$$\begin{aligned} &\{\mathcal{X}_k\}_{\text{wireless}} \triangleq \left\{\boldsymbol{x}_i \triangleq \left[\mathbf{p}_i\right]\right.; \\ &\left.\boldsymbol{f}_\ell \triangleq \left\{\boldsymbol{v}_\ell \triangleq \boldsymbol{a}_\ell, \boldsymbol{h}_\ell \triangleq \left\{\boldsymbol{e}_\ell, \boldsymbol{m}_\ell\right\}, \boldsymbol{J}_\ell\right\}; \mathrm{L}_k \triangleq \left\{\mathrm{A}_k, \mathrm{E}_k, \mathrm{M}_k, \mathrm{J}_k\right\}\right\} \end{aligned} \tag{7}$$

$$\begin{aligned} &\{\mathcal{X}_k\}_{\text{fiducial}} \triangleq \left\{\boldsymbol{x}_i \triangleq \left[\mathbf{R}_i, \mathbf{p}_i\right]\right.; \\ &\left.\boldsymbol{f}_\ell \triangleq \left\{\boldsymbol{v}_\ell \triangleq \boldsymbol{F}_\ell, \boldsymbol{h}_\ell \triangleq \left\{\boldsymbol{e}_\ell, \boldsymbol{m}_\ell\right\}, \boldsymbol{J}_\ell\right\}; \mathrm{L}_k \triangleq \left\{\mathrm{F}_k, \mathrm{E}_k, \mathrm{M}_k, \mathrm{J}_k\right\}\right\} \end{aligned} \tag{8}$$

where $\boldsymbol{e}_\ell \in \mathbb{R}^3 \in \mathrm{E}_k$ are landmarks embedded within the base structure, and $\boldsymbol{m}_\ell \in \mathbb{R}^3 \in \mathrm{M}_k$ are markers rigidly attached to visible landmarks. In proceeding formalisms, we enforce $\boldsymbol{e}_\ell$ and $\boldsymbol{m}_\ell$ as uniquely identifiable to ensure uniquely determined joint pose. For the case of anchors, only the position of the robot is considered given ToF measurements from the robot to anchors does not require knowledge of the robot′s orientation. Moreover, the inclusion of orientation as a state variable diminished system accuracy in testing, hence its exclusion. For the case of fiducial object detection and tracking, orientation of the robot is required to estimate fiducial object pose with respect to the robot′s vision system and is thus included.

3.2.2 Landmark topology

We use the term engineered landmark when formalizing landmark topologies for all $\boldsymbol{J}_\ell$:

$$\{\boldsymbol{w}_\ell^{s,s-1}\}_{\text{wireless}} = \left\{\left\{\boldsymbol{A}_\ell^{s-1}, \boldsymbol{A}_\ell^{s}\right\} \subseteq \mathrm{A}_k; \left\{\boldsymbol{E}_\ell^{s-1}, \boldsymbol{E}_\ell^{s}\right\} \subseteq \mathrm{E}_k; \left\{\boldsymbol{M}_\ell^{s-1}, \boldsymbol{M}_\ell^{s}\right\} \subseteq \mathrm{M}_k; \boldsymbol{J}_\ell \subseteq \mathrm{J}_k\right\} \tag{9}$$

$$\{\boldsymbol{w}_\ell^{s,s-1}\}_{\text{fiducial}} = \left\{\left\{\boldsymbol{F}_\ell^{s-1}, \boldsymbol{F}_\ell^{s}\right\} \subseteq \mathrm{F}_k; \left\{\boldsymbol{E}_\ell^{s-1}, \boldsymbol{E}_\ell^{s}\right\} \subseteq \mathrm{E}_k; \left\{\boldsymbol{M}_\ell^{s-1}, \boldsymbol{M}_\ell^{s}\right\} \subseteq \mathrm{M}_k; \boldsymbol{J}_\ell \subseteq \mathrm{J}_k\right\} \tag{10}$$

where $\bigcup_\ell \bigcup_s \boldsymbol{w}_\ell^{s,s-1} = \mathrm{L}_k$ and $s, s-1$ are consecutive floors in a building defining non-LOS conditions. Engineered landmark topology is defined according to:

$$\boldsymbol{A}_\ell^s = \left\{\boldsymbol{a}_\ell^s\right\} \in \mathbb{R}^{3\times m}; \boldsymbol{E}_\ell^s = \left\{\boldsymbol{e}_\ell^s\right\} \in \mathbb{R}^{3\times n}; \boldsymbol{M}_\ell^s = \left\{\boldsymbol{m}_\ell^s\right\} \in \mathbb{R}^{3\times o} \tag{11}$$

where the corresponding set of centroids $\left\{\overline{\boldsymbol{a}}_\ell^s, \overline{\boldsymbol{e}}_\ell^s, \overline{\boldsymbol{m}}_\ell^s\right\} \in \mathbb{R}^{3\times 3}$ defines the mean Euclidean distance between anchor-marker, and marker-embedded landmark sets as $\mathbf{d}\left(\overline{\boldsymbol{a}}_\ell^s, \overline{\boldsymbol{m}}_\ell^s\right)$ and $\mathbf{d}\left(\overline{\boldsymbol{m}}_\ell^s, \overline{\boldsymbol{e}}_\ell^s\right)$, respectively, for floor s. Similarly, the mean Euclidean distance between embedded landmark sets and joint position is $\mathbf{d}\left(\overline{\boldsymbol{e}}_\ell^s, \boldsymbol{J}_\ell^s(\mathbf{p}_\ell)\right)$ while the mean Euclidean distance between fiducial object position and marker landmark sets is $\mathbf{d}\left(\boldsymbol{F}_\ell^s(\mathbf{p}_\ell), \overline{\boldsymbol{m}}_\ell^s\right)$. The set of marker-embedded landmark pairings is defined by the Cartesian product of $\boldsymbol{M}_\ell^s$ and $\boldsymbol{E}_\ell^s$, where pairs define range measurements

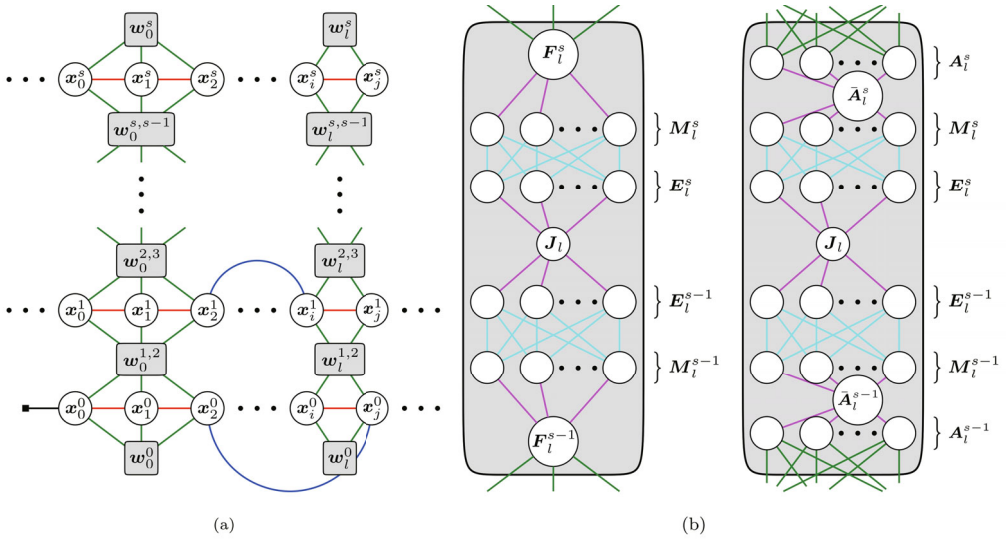

Fig. 3 (a) Factor graph representation of our proposed structural joint pose estimation framework. A prior factor (black) is assigned to the first robot state $\boldsymbol{x}_0^0$, where the subscript denotes state ID and the superscript denotes floor level (this notation is common for all state variables in this figure). Odometry factors (red) constrain consecutive robot states, loop closing factors (blue) constrain non-consecutive robot states upon re-observation of the same environmental features/landmarks, and robot-to-visible landmark factors (green) constrain the relative position between the robot and anchors (for the case of wireless sensing) and relative pose between the robot and fiducial objects (for the case of a vision system integrated with the robot). In our mathematical formalisms, we drop state superscripts and take the complete history of states as $\{\boldsymbol{x}_i\}_{i\in \mathrm{K}_k}$, where K_k is the set of keyframes. (b) Unique to our framework, we formulate visible and non-visible landmarks into topologies defining engineered landmarks $\boldsymbol{w}_\ell^{s,s-1}$, where $s, s-1$ indicate consecutive floor levels from which visible landmarks can be measured by the robot. Range-based factors (cyan) between embedded landmarks $\boldsymbol{e}_\ell \in \mathbb{R}^3 \in \boldsymbol{E}_\ell$ and marker landmarks $\boldsymbol{m}_\ell \in \mathbb{R}^3 \in \boldsymbol{M}_\ell$ form globally rigid and localizable connectivity graphs. Pose-to-position factors (pink) constrain joint states $\boldsymbol{J}_\ell \in \mathrm{SE}(3)$ with $\boldsymbol{e}_\ell$ and visible landmarks (i.e., $\boldsymbol{F}_\ell$ and $\overline{\boldsymbol{A}}_\ell$) to $\boldsymbol{m}_\ell$ and wirless anchors $\boldsymbol{a}_l$ for the case of wireless sensing. Note that $\overline{\boldsymbol{A}}_\ell$ defines the pose from which anchors and markers are commonly constrained such that $\overline{\boldsymbol{A}}_\ell(\mathbf{p}_\ell) := \overline{\boldsymbol{a}}_\ell$, where $\overline{\boldsymbol{a}}_\ell$ is the anchor set centroid.

with known correspondence and must form a globally rigid connectivity graph for the joint to be localizable from floor s. By sensing engineered landmarks from consecutive floors, repeat observations reduce drift inherent to the SLAM system via loop-closure.

3.2.3 Maximum a posteriori estimation

Unifying SLAM factors with engineered landmark factors, the MAP of $\mathcal{X}_k$ is formalized:

$$\mathcal{X}_k^* = \arg\min_{\mathcal{X}_k} \| \mathbf{r}_0 \|^2_{\mathbf{\Sigma}_0} + \underbrace{\sum_{(i,j)\in \mathrm{K}_k} \left(\| \mathbf{r}_{\mathcal{Z}_{ij}} \|^2_{\mathbf{\Sigma}_{\mathcal{Z}_{ij}}} + \sum_{\ell \in \mathrm{V}_k} \| \mathbf{r}_{\boldsymbol{x}_i, \boldsymbol{v}_\ell} \|^2_{\mathbf{\Sigma}_{\boldsymbol{x}_i, \boldsymbol{v}_\ell}} \right) + \sum_{(p,q)\in \mathrm{K}_k} \| \mathbf{r}_{\mathcal{L}_{pq}} \|^2_{\mathbf{\Sigma}_{\mathcal{L}_{pq}}}}_{\text{SLAM factors}} + \dots$$
$$\underbrace{\sum_{\ell \in \mathrm{V}_k} \sum_{\ell \in \mathrm{M}_k} \| \mathbf{r}_{\boldsymbol{v}_\ell, \boldsymbol{m}_\ell} \|^2_{\mathbf{\Sigma}_{\boldsymbol{v}_\ell, \boldsymbol{m}_\ell}} + \sum_{\ell \in \mathrm{M}_k} \sum_{\ell \in \mathrm{E}_k} \| \mathbf{r}_{\boldsymbol{m}_\ell, \boldsymbol{e}_\ell} \|^2_{\mathbf{\Sigma}_{\boldsymbol{m}_\ell, \boldsymbol{e}_\ell}} + \sum_{\ell \in \mathrm{E}_k} \sum_{\ell \in \mathrm{J}_k} \| \mathbf{r}_{\boldsymbol{e}_\ell, \boldsymbol{J}_\ell} \|^2_{\mathbf{\Sigma}_{\boldsymbol{e}_\ell, \boldsymbol{J}_\ell}}}_{\text{engineered landmark factors}} \tag{12}$$

where $\mathbf{r}_0$ is the residual error on the robot′s initial state with covariance $\mathbf{\Sigma}_0$, (i,j) are indices for consecutive keyframes, (p,q) are indices for non-consecutive keyframes as determined by loop-closure detection, $\mathbf{r}_{\mathcal{Z}_{ij}}$ is the residual error for odometry, $\mathbf{r}_{\mathcal{L}_{pq}}$ is the residual error for loop-closure detection, $\mathbf{r}_{\boldsymbol{x}_i, \boldsymbol{v}_\ell}$ is the residual error for robot-to-visible landmark measurements, $\mathbf{r}_{\boldsymbol{v}_\ell, \boldsymbol{m}_\ell}$ is the residual error for visible landmark-to-marker measurements, $\mathbf{r}_{\boldsymbol{m}_\ell, \boldsymbol{e}_\ell}$ is the residual error for range measurements between marker-embedded landmark pairs, and $\mathbf{r}_{\boldsymbol{e}_\ell, \boldsymbol{J}_\ell}$ is the residual error for embedded-to-joint landmark geometric constraints. For the case of wireless sensing, SLAM factor residuals are formalized in Eqs. (13) through (16):

$$\mathbf{r}_0 = {}^{W}_{B_0}\mathbf{p} - {}^{W}_{B_0}\tilde{\mathbf{p}} \tag{13}$$

$$\mathbf{r}_{\mathcal{Z}_{ij}} = ({}^{W}_{B_i}\mathbf{p} - {}^{W}_{B_j}\mathbf{p}) - {}^{B_j}_{B_i}\tilde{\mathbf{p}} \tag{14}$$

$$\mathbf{r}_{\mathcal{L}_{pq}} = ({}^{W}_{B_p}\mathbf{p} - {}^{W}_{B_q}\mathbf{p}) - {}^{B_q}_{B_p}\tilde{\mathbf{p}} \tag{15}$$

$$\mathbf{r}_{\boldsymbol{x}_i, \boldsymbol{a}_\ell} = \mathbf{d}({}^{W}_{a_l}\mathbf{p},\ {}^{W}_{B_i}\mathbf{p}) - \mathbf{d}({}^{W}_{a_l}\tilde{\mathbf{p}},\ {}^{W}_{B_i}\tilde{\mathbf{p}}) \tag{16}$$

where $\widetilde{(\cdot)}$ indicates a measured quantity, ${}^{A}_{\mathrm{B}}\mathbf{p}$ is the measured position in frame B with respect to frame A, W defines the world coordinate frame, and B defines the body frame of the robot from which all robot-to-visible landmark measurements are assumed to be captured. Note that in practice, measurements taken with respect to sensor frame S can be transformed into B using extrinsic calibrations. SLAM factor residuals considering fiducial object detection and tracking are formalized in Eqs. (17) through (20):

$$\mathbf{r}_0 = \mathbf{T}_{WB_0} \ominus \tilde{\mathbf{T}}_{WB_0} \tag{17}$$

$$\mathbf{r}_{\mathcal{Z}_{ij}} = \Phi\left(\tilde{\mathbf{T}}^{-1}_{WB_i}\tilde{\mathbf{T}}_{WB_j}, \mathbf{T}^{-1}_{WB_i}\mathbf{T}_{WB_j}\right) \tag{18}$$

$$\mathbf{r}_{\mathcal{L}_{pq}} = \Phi\left(\tilde{\mathbf{T}}^{-1}_{WB_p}\tilde{\mathbf{T}}_{WB_q}, \mathbf{T}^{-1}_{WB_p}\mathbf{T}_{WB_q}\right) \tag{19}$$

$$\mathbf{r}_{\boldsymbol{x}_i, \boldsymbol{F}_\ell} = \Phi\left(\tilde{\mathbf{T}}^{-1}_{WB_i}\tilde{\mathbf{T}}_{WF_l}, \mathbf{T}^{-1}_{WB_i}\mathbf{T}_{WF_l}\right) \tag{20}$$

where $\mathbf{T}_{AB} = \begin{bmatrix}\mathbf{R} \mid \mathbf{p}\end{bmatrix} \in \mathrm{SE}(3)$ is the homogeneous transformation matrix expressing the coordinate transform from A to B in A such that $\mathbf{T}_1 \ominus \mathbf{T}_2 = \mathbf{T}_1^{-1}\mathbf{T}_2$ and $\mathbf{T}_1 \otimes \mathbf{T}_2 = \mathbf{T}_1\mathbf{T}_2$, and $\Phi(\cdot)$ is the lifting operator defined in (Forster *et al.*, 2016). Engineered landmark factor residuals are:

$$\mathbf{r}_{\bar{\boldsymbol{A}}_\ell, \boldsymbol{a}_\ell} = \left(\mathbf{T}^{-1}_{W\bar{A}_l} \otimes {}^{W}_{a_l}\mathbf{p}\right) \ominus {}^{\bar{A}_l}_{a_l}\tilde{\mathbf{p}} \tag{21}$$

$$\mathbf{r}_{\bar{\boldsymbol{A}}_\ell, \boldsymbol{m}_\ell} = \left(\mathbf{T}^{-1}_{W\bar{A}_l} \otimes {}^{W}_{m_l}\mathbf{p}\right) \ominus {}^{\bar{A}_l}_{m_l}\tilde{\mathbf{p}} \tag{22}$$

$$\mathbf{r}_{\boldsymbol{F}_\ell, \boldsymbol{m}_\ell} = \left(\mathbf{T}^{-1}_{WF_l} \otimes {}^{W}_{m_l}\mathbf{p}\right) \ominus {}^{F_l}_{m_l}\tilde{\mathbf{p}} \tag{23}$$

$$\mathbf{r}_{\boldsymbol{m}_\ell, \boldsymbol{e}_\ell} = \mathbf{d}({}^{W}_{m_l}\mathbf{p},\ {}^{W}_{e_l}\mathbf{p}) - \mathbf{d}({}^{W}_{m_l}\tilde{\mathbf{p}},\ {}^{W}_{e_l}\tilde{\mathbf{p}}) \tag{24}$$

$$\mathbf{r}_{\boldsymbol{e}_\ell, \boldsymbol{J}_\ell} = \left(\mathbf{T}^{-1}_{WJ_l} \otimes {}^{W}_{e_l}\mathbf{p}\right) \ominus {}^{J_l}_{e_l}\tilde{\mathbf{p}} \tag{25}$$

where $\left\{\bar{\boldsymbol{A}}_\ell \in \mathrm{SE}(3) \mid \bar{\boldsymbol{A}}_\ell(\mathbf{p}_\ell) \coloneqq \bar{\boldsymbol{a}}_\ell\right\}$ defines the pose from which anchors and markers are commonly constrained and is made possible by the use of rigid-body connections between anchors and markers.

4 Experimental evaluation

To evaluate our framework, we develop a custom simulator that allows for the construction and optimization of our factor graph with controlled measurements. Measurements are derived from ground-truth quantities perturbed by zero-bias Gaussian noise such that:

$$\tilde{\mathbf{R}} = \mathbf{R}\mathrm{Exp}(\boldsymbol{\epsilon});\ \tilde{\mathbf{p}} = \mathbf{p} + \boldsymbol{\epsilon} \tag{26}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ and $\mathrm{Exp}(\cdot)$ is the exponential map operator defined in Forster *et al.* (2016). This allows for precise system noise parameterization, generalizing our results to SLAM systems and sensors meeting specific accuracy requirements. In doing so, residual joint pose errors are attributed to system noise and optimization convergence only. To demonstrate our framework in a future laboratory experiment, Fig. 4 illustrates a potential

set-up where methods addressing non-Gaussian noise (Huang *et al.*, 2022a, 2022b), and outliers (Yang *et al.*, 2020a; Shi *et al.*, 2021; Yang and Carlone, 2023) can be applied to real-world measurements. To arrive at residual joint pose error metrics and IDR results summarized in Section 4.4, we detail our simulation environment and modelling assumptions next.

4.1 Simulation environment

We simulate the robotic inspection of a two-bay, 19-story building illustrated in Fig. 5(a), where engineered landmarks exist along the perimeter of the building to coincide with beam-column connections in the SFRS. We carry out inspections on the undeformed structure and note that landmark displacements can be equally determined pre- and post-earthquake given LOS conditions remain between visible landmarks and the robot. Inspection begins with the robot initializing from an arbitrary frame $\mathcal{F}_{map}$ from which measurements are globally referenced. For convenience, we assume the world frame $\mathcal{F}_W$ coincides with $\mathcal{F}_{map}$, simplifying error calculations as ground truth quantities are referenced from $\mathcal{F}_W$. In practice, the initialization method by Watanabe (2018) may be implemented for real-world deployment to ensure pre- and post-earthquake maps are referenced to a common $\mathcal{F}_{map}$. For our study, we model uncertainty on the robot's initialization for these methods via the prior on $\boldsymbol{x}_0$. Alternatively, post-earthquake inspections may be carried out where estimated joint states are compared to the as-built structural model of the building.

To ensure all visible landmarks are measured by the robot during inspection, Fig. 5(b) illustrates the assignment of waypoints to suite the robot's wireless sensing or vision system. Starting from floor $s=0$, the robot navigates to each waypoint, completing a single loop prior to transitioning to the floor above. By representing the ground truth trajectory in continuous time using a second-order, b-spline curve in $\mathrm{SE}(3)$, we sample ground truth poses at a key-frame rate of 0.5 s assuming a constant robot velocity of 0.5 m/s. We limit the maximum wireless range (between anchors and the robot) to the bay width w and accordingly produce coverage envelopes that divide the total number of keyframes evenly between all engineered landmarks so they are equally determined. For the case of fiducial object detection and tracking, we again limit the vision system's field of view to w and assume a linear noise model irrespective of incidence angle:

$$\begin{aligned}\mu_{\boldsymbol{x}_i,\boldsymbol{F}_\ell} &= \frac{\mu_{\pi(\boldsymbol{F}_\ell)}}{f}\mathbf{d}\left(\boldsymbol{x}_i,\boldsymbol{F}_\ell\left(\mathbf{p}_\ell\right)\right);\\ \sigma_{\boldsymbol{x}_i,\boldsymbol{F}_\ell} &= \frac{\sigma_{\pi(\boldsymbol{F}_\ell)}}{f}\mathbf{d}\left(\boldsymbol{x}_i,\boldsymbol{F}_\ell\left(\mathbf{p}_\ell\right)\right)\end{aligned} \tag{27}$$

where $\mu_{\pi(\boldsymbol{F}_\ell)}$ and $\sigma_{\pi(\boldsymbol{F}_\ell)}$ define the mean and standard deviation of visual reprojection errors (in pixels) when detecting fiducial objects (Kallwies *et al.*, 2020) and f is camera focal length. Lastly, Fig. 5(c) illustrates the chosen topologies for this study, where $\mathbf{d}\left(\overline{\boldsymbol{e}}_\ell,\boldsymbol{J}_\ell\left(\mathbf{p}_\ell\right)\right)$ and $\mathbf{d}\left(\overline{\boldsymbol{e}}_\ell,\overline{\boldsymbol{m}}_\ell\right)$ parameterize uncertainty inherent to engineered landmark factors as described in Section 4.4.

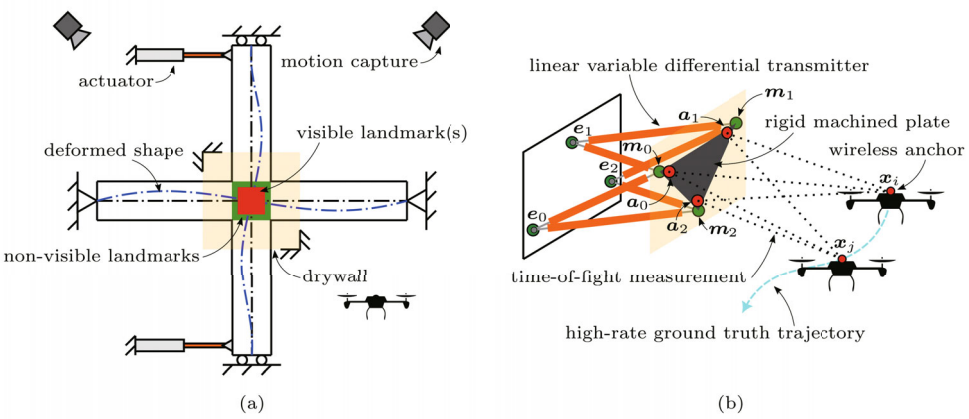

Fig. 4 (a) 2D view of a potential laboratory experiment that may be established to validate our framework. A scaled seismic force-resisting system (SFRS) is actuated to simulate beam-column residual joint displacement under various service level and maximum considered earthquakes. A motion capture system is used to capture the ground truth position/pose of the robot, while non-line-of-sight between visible and non-visible landmarks is introduced by a piece of drywall that is isolated from the scaled SFRS. Framework assumptions, including negligible panel zone shearing and visible landmarks remaining visible to the robot post-earthquake (as discussed in Section 4.3), can be checked using this set-up. Note the ground truth shape of the scaled SFRS may be obtained using dense structure from motion or a terrestrial laser scanner. (b) A particular instantiation of the engineered landmark may involve three embedded landmarks, three markers, and three anchors. Embedded landmarks may connect to markers via linear variable differential transmitters, while anchors can be sensed by a single anchor integrated with the robot. The high-rate ground truth trajectory of the robot (obtained by the motion capture system) can be sampled and perturbed to represent the current state-of-the-art SLAM system. A rigid machined plate may connect markers and anchors and can be anchored into the drywall

4.2 Finite element modelling

We evaluate how our proposed method can predict inter-story drift by constructing a finite element (FE) model of a tall, slender building. For this study, a steel moment-resisting frame (MRF) model was obtained from a database of seismic designs containing over 100 different geometric configurations, each designed under site-specific responses (Guan *et al.*, 2021). The seismic design database contains nonlinear 2D FE OpenSees (Mazzoni *et al.*, 2006) models for each design, allowing time-history analyses (under different ground motion histories) to simulate realistic joint displacements. We chose the most slender building design, containing 19 storeys and two bays, with $h = 13$ ft (3.96 m) floor height per level and $w = 20$ ft (6.092 m) bay width. The model is symmetric in the longitudinal and transverse directions, which allows us to assume that the 2D displacements of the model can be generalized to an equivalent 3D model (with the same geometry) in the orthogonal directions.

Nonlinear effects are considered by including

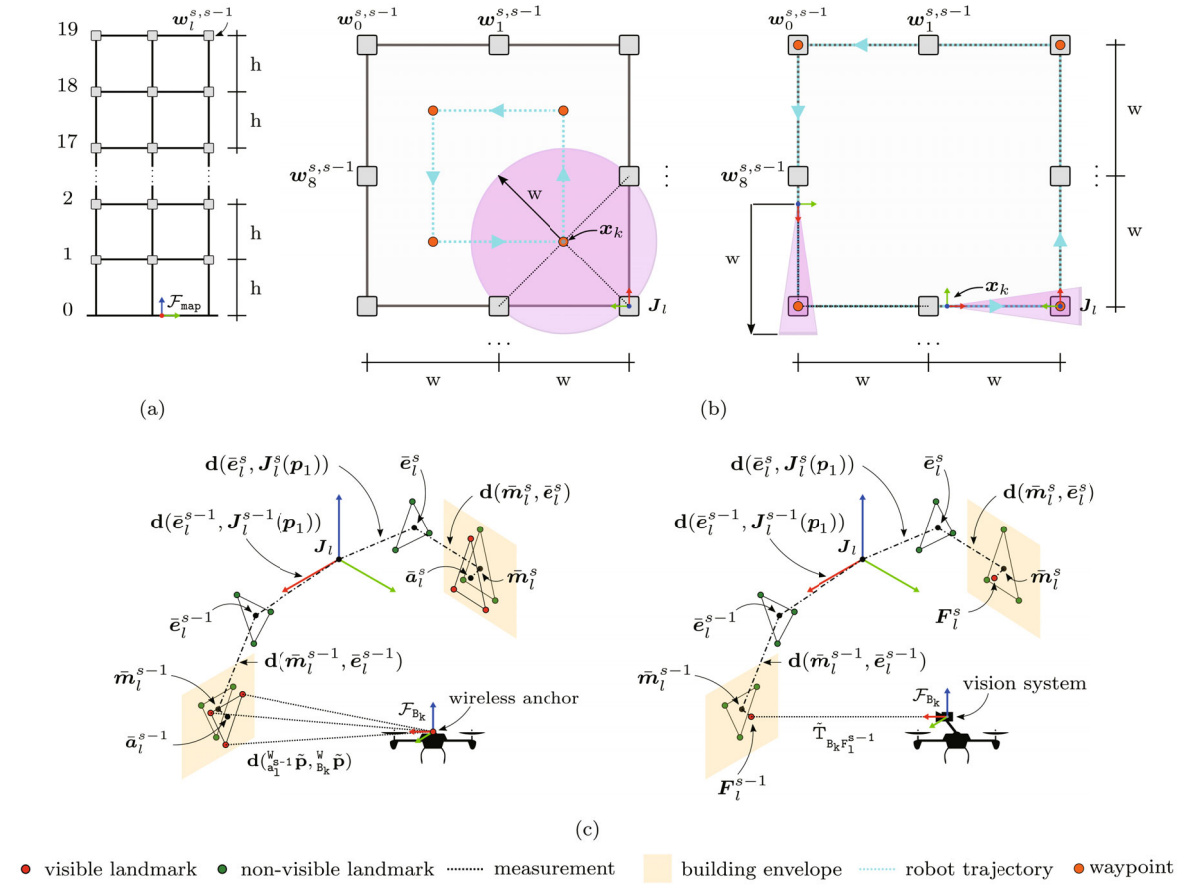

Fig. 5 **(a) Side-profile of the two bay, 19 storey nonlinear structural model used to evaluate our framework. $\mathcal{F}_{map}$ defines the frame of a landmark (ex. fiducial object) from which the SLAM system initializes. For this study, we model this initialization via uncertainty on the first robot state via a prior factor and assume $\mathcal{F}_{map}$ coincides with the same global coordinate system as the structural model (i.e., the world coordinate frame $\mathcal{F}_W$). (b) Top-down view of floor s. Waypoints (orange circles) define the robot's ground truth trajectory (cyan dashed line), where the robot navigates to each waypoint to complete the inspection of all engineered landmarks per floor. Pink envelopes illustrate the robot's measurement field of view (FOV), where the robot's trajectory is adjusted to suit its sensing modality. For the case of wireless sensing (left), we envision a single wireless anchor node integrated with the robot to measure the relative position of the robot with respect to wireless sensor networks integrated with engineered landmarks and assume measurement uncertainty remains uniform within the FOV. For the case of fiducial object detection and tracking (right), we control for the robot's trajectory to ensure the incidence angle between fiducial objects and the robot's vision system minimally affects accuracy and thus assume measurement uncertainty to reduce linearly as the robot approaches engineered landmarks in our simulations. (c) Engineered landmark topologies chosen for our study. Here, $\mathbf{d}(\bar{\boldsymbol{e}}_\ell, \boldsymbol{J}_\ell(\mathbf{p}_\ell))$ defines the Euclidean distance between joint position $\boldsymbol{J}_\ell(\mathbf{p}_\ell)$ and embedded landmark set centroids $\bar{\boldsymbol{e}}_\ell$ and $\mathbf{d}(\bar{\boldsymbol{m}}_\ell, \bar{\boldsymbol{e}}_\ell)$ defines the Euclidean distance between $\bar{\boldsymbol{e}}_\ell$ and marker landmark set centroids $\bar{\boldsymbol{m}}_\ell$. For this study, we take $|\boldsymbol{E}_\ell^{s-1}| = |\boldsymbol{E}_\ell^{s}| = |\boldsymbol{M}_\ell^{s-1}| = |\boldsymbol{M}_\ell^{s}| = |\boldsymbol{A}_\ell^{s-1}| = |\boldsymbol{A}_\ell^{s}| = 3$ and assume fiducial object pose $\boldsymbol{F}_\ell$ and the pose defined by anchors $\bar{\boldsymbol{A}}_\ell$ (where $\bar{\boldsymbol{A}}_\ell(\mathbf{p}_\ell)$ equates to anchor set centroids $\bar{\boldsymbol{a}}_\ell$) are constrained with $\bar{\boldsymbol{m}}_\ell$ via rigid connections that penetrate the building envelope**

geometric and material nonlinearity. For geometric nonlinearity, a P-Delta analysis is performed to simulate the influence of gravity loads on the lateral stiffness of the building. Uniformly distributed gravity loads are applied at each level of the building, and a stability analysis is performed before and during ground motion. To simulate joint displacement during an earthquake, a concentrated plasticity model is used to simulate the nonlinear interaction in the beam-column region using inelastic hinges at both ends of each column, which also account for shear yielding in the panel zone. For the dynamic analysis, the Rayleigh damping ratio was set to 2%. A detailed summary of loading, section properties, and nonlinear material parameters used for this study can be found in the seismic design database, model 1539.

The FE model is evaluated under service-level earthquake (SLE) and maximum considered earthquake (MCE) hazard levels, each containing 38 and 40 records, respectively, from the seismic designs database. 2D displacements from nonlinear time history analyses are represented in 3D where ground motions remain restricted to the 3D FE model′s longitudinal direction. This assumption is made given the 2D models are designed to satisfy SLE and MCE limit states; thus, combining their responses in orthogonal directions would over-estimate allowable IDR. Residual joint displacements from the time history analyses establish the ground-truth from which maximum average IDRs are calculated in Section 4.4.

4.3 Framework modeling assumptions

We assume the following when evaluating our framework:

(1) Embedded landmarks are secured to structural joints such that panel zone shear yielding is negligible (i.e., embedded landmarks undergo the same coordinate transformation as structural joints during the earthquake event).

(2) Visible landmarks remain visible to the robot post-earthquake, with rigid connections between visible and non-visible landmarks remaining secured within the building envelope.

(3) $\mathcal{F}_{map}$ remains globally consistent pre- and post-earthquake.

Further assumptions are made for the simulation conducted in this study:

(1) Measurement errors are precisely modelled by their associated covariance matrices when solving Eq. (12)

(2) Odometry measurements implicitly capture loop-closures identified within each floor, with error uniformly distributed across the set of measured positions (for the case of wireless sensing) and poses (for the case of fiducial object detection and tracking). As such, we do not explicitly account for loop-closing measurements $\mathcal{L}_{pq}$.

(3) Robot-to-visible landmark measurements occur at keyframe rate, with zero-drift time synchronization between all sensors. In practice, asynchronous ToF wireless measurements are rectified using the methods of (Nguyen *et al.*, 2021b) and would be implemented when deploying our system on hardware.

(4) MAP estimation is modelled as an offline process.

(5) Measurements taken by the robot at keyframe $k \in \mathrm{K}_k$ are expressed in its body frame $\mathcal{F}_{B_k}$ using ground truth calibrations.

4.4 Experimental evaluation

Experiments applying the Monte Carlo method characterize our framework′s performance as a function of system noise. To assess the applicability of our approach, the following noise parameters are established.

4.4.1 SLAM factors

The following odometry parameterization is consistent with values reported by Ramezani *et al.* (2022) for their state-of-the-art lidar-inertial SLAM system on the Hilti dataset (Helmberger *et al.*, 2022), which represents a best-case scenario for robotic infrastructure inspection:

$$\mu_{\mathcal{Z}_{ij}} = \left\{ \underbrace{\{0.9\ \text{cm}\}_{i=1}^{3}}_{\text{wireless sensing}}, \underbrace{\{\{0.0^\circ\}_{i=1}^{3}, \{0.9\ \text{cm}\}_{i=1}^{3}\}}_{\text{fiducial object detection and tracking}} \right\};$$

$$\boldsymbol{\Sigma}_{\mathcal{Z}_{ij}} = \left\{ \underbrace{\text{diag}(\{0.5\ \text{cm}\}_{i=1}^{3})^2}_{\text{wireless sensing}}, \underbrace{\text{diag}(\{0.5^\circ\}_{i=1}^{3}, \{0.5\ \text{cm}\}_{i=1}^{3})^2}_{\text{fiducial object detection and tracking}} \right\} \tag{28}$$

For the case of wireless sensing, we parameterize robot-to-anchor measurements according to the theoretical upper bound for UWB (Oppermann *et al.*, 2004) and low cost UWB systems that are commercially available and economically viable for small scale infrastructure (Nguyen *et al.*, 2021c):

$$\mu_{\boldsymbol{x}_i, \boldsymbol{a}_\ell} = 0;\ \boldsymbol{\Sigma}_{\boldsymbol{x}_i, \boldsymbol{a}_\ell} = \left\{ \underbrace{1\ \text{cm}}_{\text{theoretical}}\ ,\ \underbrace{5^2\text{cm}}_{\text{commercial}} \right\} \tag{29}$$

For the case of fiducial object detection and tracking, we parameterize robot-to-fiducial object measurements according to Eq. (27) and set $\mu_{\pi(\boldsymbol{F}_\ell)} = 0.017$ px, $\sigma_{\pi(\boldsymbol{F}_\ell)} = 0.01$ px, and $f = 1130$ px (as per Kallwies *et al.*, (2020)) to approximate the state-of-the-art with controlled robot navigation:

$$\boldsymbol{\Sigma}_{\boldsymbol{x}_i, \boldsymbol{F}_\ell} = \text{diag}(\{\sigma_{\boldsymbol{x}_i, \boldsymbol{F}_\ell}\}_{i=1}^{3})^2 \tag{30}$$

4.4.2 Engineered landmark factors

For generality, we parameterize engineered landmark factors as a function of mean Euclidean distance according to:

$$\mu_{\boldsymbol{m}_\ell,\boldsymbol{e}_\ell} = 0 \text{ cm};\ \boldsymbol{\Sigma}_{\boldsymbol{m}_\ell,\boldsymbol{e}_\ell} = \left\{ \underbrace{0.001\%\mathbf{d}^2_{\bar{\boldsymbol{m}}_\ell,\bar{\boldsymbol{e}}_\ell}}_{\text{theoretical LVDT}}, \underbrace{1\%\mathbf{d}^2_{\bar{\boldsymbol{m}}_\ell,\bar{\boldsymbol{e}}_\ell}}_{\text{practical LVDT}} \right\}; \quad (31)$$

$$\mu_{\boldsymbol{e}_\ell,\boldsymbol{J}_\ell} = \{0 \text{ cm}\}_{i=1}^3;$$

$$\boldsymbol{\Sigma}_{\boldsymbol{e}_\ell,\boldsymbol{J}_\ell} = \left\{ \underbrace{\text{diag}(\{1\%\boldsymbol{d}_{\bar{\boldsymbol{e}}_\ell,\boldsymbol{J}_\ell}\}_{i=1}^3)^2}_{\text{precise installation}}, \underbrace{\text{diag}(\{10\%\boldsymbol{d}_{\bar{\boldsymbol{e}}_\ell,\boldsymbol{J}_\ell}\}_{i=1}^3)^2}_{\text{course installation}} \right\} \quad (32)$$

For marker-to-embedded landmark factors, percentages are adjusted to align theoretical and practical LVDT range measurements, while embedded-to-joint percentages reflect precise and course installations respectively. In our experiments, all engineered landmarks are configured such that:

(1) The cardinality of anchor, marker, and embedded landmark sets is three: $|\boldsymbol{A}_\ell| = |\boldsymbol{M}_\ell| = |\boldsymbol{E}_\ell| = 3$

(2) Mean Euclidean distances are:

$\mathbf{d}_{\bar{\boldsymbol{m}}_\ell,\bar{\boldsymbol{e}}_\ell} := \mathbf{d}(\bar{\boldsymbol{m}}_\ell,\bar{\boldsymbol{e}}_\ell)^{s-1} = \mathbf{d}(\bar{\boldsymbol{m}}_\ell,\bar{\boldsymbol{e}}_\ell)^s$,

$\mathbf{d}_{\bar{\boldsymbol{e}}_\ell,\boldsymbol{J}_\ell} := \mathbf{d}(\bar{\boldsymbol{e}}_\ell,\boldsymbol{J}_\ell(\mathbf{p}_\ell))^{s-1} = \mathbf{d}(\bar{\boldsymbol{e}}_\ell,\boldsymbol{J}_\ell(\mathbf{p}_\ell))^s$

(3) Anchor, fiducial object, and marker set centroids coincide: $\bar{\boldsymbol{a}}_\ell^{s-1} = \boldsymbol{F}_\ell^{s-1}(\mathbf{p}_\ell) = \bar{\boldsymbol{m}}_\ell^{s-1}$, $\bar{\boldsymbol{a}}_\ell^s = \boldsymbol{F}_\ell^s(\mathbf{p}_\ell) = \bar{\boldsymbol{m}}_\ell^s$

(4) Anchors and markers are precisely secured to a common rigid plate such that: $\boldsymbol{\Sigma}_{\bar{\boldsymbol{A}}_\ell,\boldsymbol{a}_\ell} = \boldsymbol{\Sigma}_{\bar{\boldsymbol{A}}_\ell,\boldsymbol{m}_\ell} = \text{diag}(\{0.1 \text{ cm}\}_{i=1}^3)^2$

4.4.3 Results

With the above parameterization, Table 1 provides summary statistics for residual joint translation and rotation errors for $n_{\text{realizations}} = 100$ Monte Carlo realizations according to:

$$\mathbf{r}_{\text{translation}} = \| \boldsymbol{J}_\ell(\mathbf{p}_\ell) - \tilde{\boldsymbol{J}}_\ell(\mathbf{p}_\ell) \| \quad (33)$$

$$\mathbf{r}_{\text{rotation}} = \| \angle\left(\boldsymbol{J}_\ell(\mathbf{R}_\ell)\tilde{\boldsymbol{J}}_\ell(\mathbf{R}_\ell^\top)\right) \| \quad (34)$$

where $\angle(\cdot)$ expresses rotations in Euler angles. Residual joint pose error distributions illustrated in Figs. 6(a) and 6(b) show the importance of precise embedded landmark installation, where Cases B and D prove unreliable in rotation estimation for all cases. For precise embedded landmark installation, fiducial object detection

Table 1 Monte Carlo simulation summary statistics (residual translation error (cm), residual rotation error (°), maximum average inter-story drift ratio (%) error) under a sensitivity of system noise*

	Theoretical LVDT		Practical LVDT	
	Case A: Precise Installation	Case B: Course Installation	Case C: Precise Installation	Case D: Course Installation
	Case 1: Theoretical wireless sensing			
mean	(0.90, 0.19, 0.02)	(1.76, 1.42, 0.07)	(0.90, 0.19, 0.02)	(1.76, 1.42, 0.07)
median	(0.81, 0.17, 0.02)	(1.59, 1.26, 0.06)	(0.86, 0.18, 0.02)	(1.61, 1.26, 0.05)
min	(0.02, 0.00, 0.00)	(0.03, 0.03, 0.00)	(0.02, 0.01, 0.00)	(0.02, 0.03, 0.00)
max	(4.02, 0.95, 0.11)	(6.97, 9.59, 0.36)	(3.48, **0.93**, 0.11)	(7.42, 9.84, 0.32)
std. dev.	(0.46, **0.10**, 0.02)	(0.96, 0.83, 0.06)	(0.45, **0.10**, 0.02)	(0.95, 0.83, 0.05)
RMSE	(1.01, 0.22, 0.03)	(2.01, 1.65, 0.09)	(1.02, 0.22, 0.03)	(2.00, 1.64, 0.08)
	Case 2: Commercial wireless sensing			
mean	(1.78, 0.21, 0.06)	(2.38, 1.65, 0.09)	(1.77, 0.21, 0.06)	(2.39, 1.66, 0.09)
median	(1.60, 0.18, 0.05)	(2.12, 1.46, 0.07)	(1.59, 0.19, 0.05)	(2.14, 1.46, 0.07)
min	(0.07, 0.00, 0.00)	(0.04, 0.01, 0.00)	(0.02, 0.00, 0.00)	(0.08, 0.05, 0.00)
max	(23.74, 1.68, 0.80)	(31.93, 28.65, 0.70)	(17.88, 1.06, 0.32)	(29.19, 38.16, 0.82)
std. dev.	(1.09, 0.13, 0.05)	(1.40, 1.08, 0.07)	(1.08, 0.12, 0.04)	(1.43, 1.23, 0.08)
RMSE	(2.08, 0.24, 0.07)	(2.76, 1.97, 0.12)	(2.08, 0.24, 0.07)	(2.78, 2.07, 0.12)
	Case 3: Fiducial object detection and tracking			
mean	(**0.43**, **0.16**, **0.01**)	(0.65, 1.34, 0.06)	(0.54, **0.16**, **0.01**)	(0.64, 1.33, 0.06)
median	(**0.31**, **0.14**, **0.01**)	(0.58, 1.18, 0.05)	(0.34, **0.14**, **0.01**)	(0.57, 1.18, 0.05)
min	(**0.01**, 0.00, 0.00)	(**0.01**, 0.00, 0.00)	(**0.01**, 0.00, 0.00)	(0.01, 0.02, 0.00)
max	(**3.53**, 1.64, **0.07**)	(5.38, 8.42, 0.24)	(5.31, 1.76, 0.08)	(4.12, 7.60, 0.28)
std. dev.	(**0.40**, **0.10**, **0.01**)	(0.36, 0.78, 0.04)	(0.59, 0.11, **0.01**)	(0.36, 0.77, 0.04)
RMSE	(**0.58**, **0.19**, **0.02**)	(0.74, 1.55, 0.07)	(0.80, **0.19**, **0.02**)	(0.73, 1.54, 0.07)

* The number of Monte Carlo realizations is $n_{\text{realizations}} = 100$. Bold values represent best performance among considered cases. See Section 4.4 for corresponding parameter values.

and tracking (i.e., Case 3) and theoretical wireless sensing (i.e., Case 1) show similar performance, while commercial wireless sensing (i.e., Case 2) introduces a significant number of outliers beyond centimeter-level translation accuracy. These observations are further supported by statistics summarized in Table 1, where fiducial object detection and tracking excels across most metrics while theoretical wireless sensing closely follows. Based on maximum residual joint translation error, commercial wireless sensing proves unreliable in joint translation estimation and is thus unsafe for post-earthquake damage assessment.

As an initial assessment of applicability, we evaluate the reliability and repeatability of IDR estimates by calculating translation error (in the equivalent longitudinal direction of the 3D FE models) between engineered landmarks on adjacent floors. For all cases and Monte Carlo realizations, we average the nine discrete translation errors (per floor) to represent average IDR error, and take the maximum value. Figure 6(c) illustrates maximum average IDR distributions for the previously summarized cases, with summary statistics included in Table 1. Again, fiducial object detection and tracking and theoretical wireless sensing exhibit similar performance with significant accuracy benefits over commercial wireless sensing.

We further evaluate applicability by considering $\mathrm{SLE}_{\delta} = 0.5\%$ and $\mathrm{MCE}_{\delta} = 1.0\%$ IDR limit states, which have been adopted from the Tall Building Initiative (TBI) (Pacific Earthquake Engineering Reasearch Centre (PEER), 2017). These limit states represent the maximum allowable IDR at extreme points (for each floor) to protect against permanent lateral deformation, with more stringent requirements for SLE to prevent non-structural component damage. For all SLE and MCE earthquake responses, we 1) take the maximum IDR, 2) bound results with maximum average IDR errors over all Monte Carlo realizations, 3) and normalize these errors by SLE_{δ} and MCE_{δ} as shown in Fig. 7.

Similar to maximum average IDR error distributions, results show course embedded landmark installation limits the reliability of our framework, with [0.48, 0.85] and [0.23, 0.43] min-max SLE and MCE utilization errors yielding post-earthquake assessment unsafe for fiducial object detection and tracking and theoretical wireless sensing. Given precise installation, fiducial object detection and tracking remains within 10% tolerance of MCE_{δ} for reliable and repeatable MCE post-earthquake

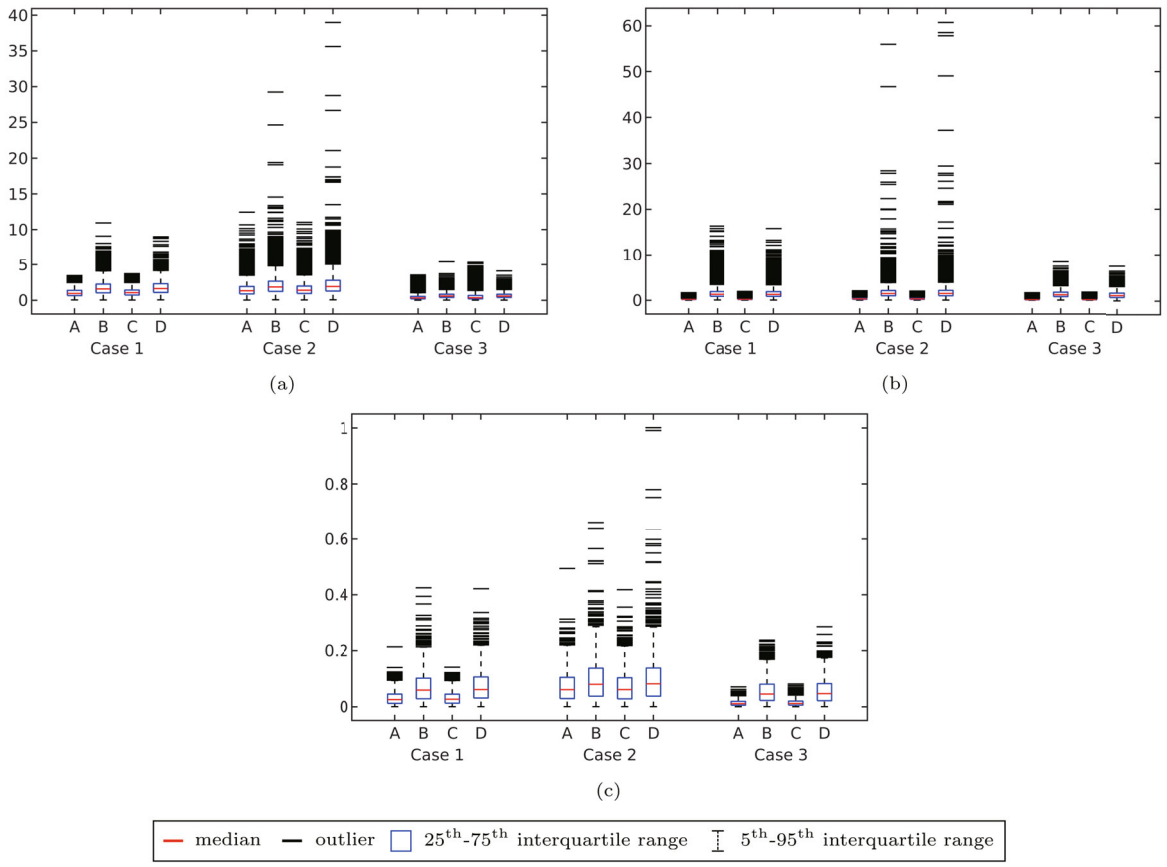

Fig. 6 (a) Residual translation error (cm) (b) residual rotation error (°) and (c) maximum average inter-story drift ratio error (%) box plots for the following cases: A: (theoretical LVDT, precise installation), B: (theoretical LVDT, course installation), C: (practical LVDT, precise installation), D: (practical LVDT, course installation), Case 1: theoretical wireless sensing, Case 2: commercial wireless sensing, Case 3: fiducial object detection and tracking. See Section 4.4 for corresponding parameter values

assessment, while all other cases exceed this threshold. For SLE post-earthquake assessment, the 10% tolerance is exceeded in all cases. As a course estimate of SLE and MCE IDR limit state utilization, Fig. 7 suggests fiducial object detection and tracking and theoretical wireless sensing is adequate for the particular FE model chosen in this study.

5 Conclusion

In this paper, we present a novel graph-based state estimation framework that enables 6-DoF structural joint localization in non-line-of-sight conditions. By unifying simultaneous localization and mapping and engineered landmark factors within a common graph, we solve the maximum a posterior estimate of robot, landmark, and joint states via nonlinear least squares optimization. We asses the applicability of our framework via the Monte Carlo method in simulation, where state uncertainty is modelled after state-of-the-art simultaneous localization and mapping, theoretical and commercial wireless sensing, state-of-the-art fiducial object detection and tracking, and precise and course installation errors. Over a suite of realizations and case-specific noise

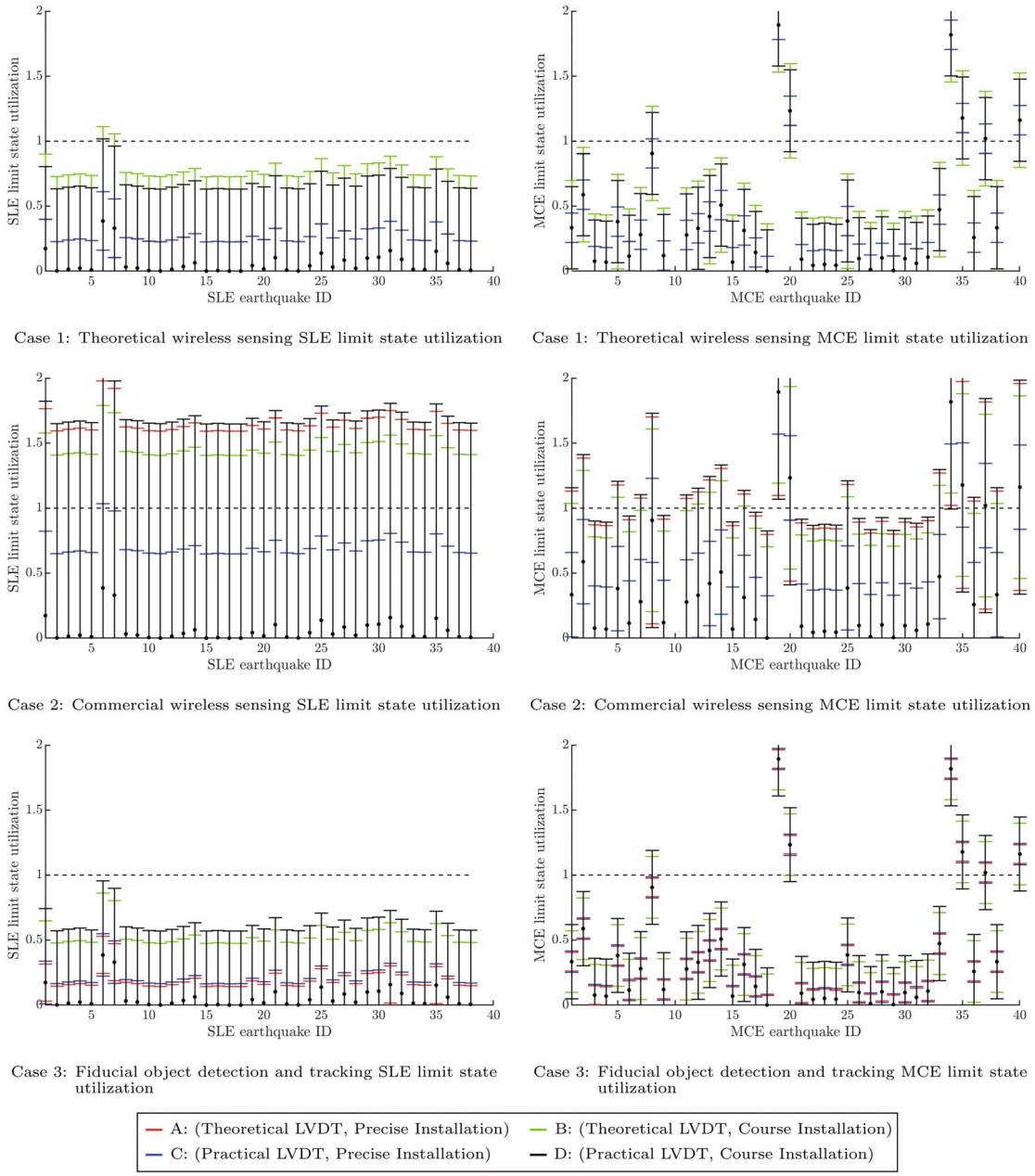

Fig. 7 Maximum service-level earthquake (SLE) and maximum considered earthquake (MCE) inter-story drift ratios normalized to $\mathrm{SLE}_{\delta} = 0.5\%$ and $\mathrm{MCE}_{\delta} = 1.0\%$ IDR limit states, respectively. SLE and MCE limit states have been adopted from the Tall Building Initiative (TBI) (Pacific Earthquake Engineering Research Centre (PEER), 2017), where measured limit states below 1.00 (dashed black line) utilization indicate adequate performance. Error bounds (indicated in the legend) represent the maximum average IDR error over all Monte Carlo realizations

parameters, we quantify residual joint pose error distributions, maximum average inter-story drift ratios, and related summary statistics for a 19-story nonlinear structural model. Experimental results show fiducial object detection and tracking has the potential to offer accurate and repeatable inter-story drift estimates given precise engineered landmark topology as a means of global damage assessment. Following this study, we aim to validate the real-world performance of our proposed framework in laboratory testing.

Acknowledgment

This work was supported by the Natural Sciences and Engineering Research Council of Canada through their Research Tools and Instruments and Fellowship programs, and the Graduate Fellowship program by the University of California, Los Angeles.

References

Aiello GR and Rogerson GD (2003), "Ultra-Wideband Wireless Systems," *IEEE Microwave Magazine*, **4**(2): 36–47.

Applied Technology Council (2005), *Field Manual : Postearthquake Safety Evaluation of Buildings*, *2nd ed*., Applied Technology Council, Redwood City, CA, USA.

ASCE/SEI 43-05 (2005), *Seismic Design Criteria for Structures, Systems, and Components in Nuclear Facilities*, American Society of Civil Engineers, Reston, VA, USA.

ASCE/SEI 7-16 (2016), *Minimum Design Loads and Associated Criteria for Buildings and Other Structures*, American Society of Civil Engineers, Reston, VA, USA.

Cao Y and Beltrame G (2021), "VIR-SLAM: Visual, Inertial, and Ranging Slam for Single and Multi-Robot Systems," *Autonomous Robots*, **45**: 905–917.

Çelebi M (2005), "Structural Monitoring Arrays—Past, Present and Future," in P Gűlkan and JG Anderson, Editors, *Directions in Strong Motion Instrumentation*, 157–179.

Chepuri SP, Leus G and van der Veen A-J (2014), "Rigid Body Localization Using Sensor Networks," *IEEE Transactions on Signal Processing*, **62**(18): 4911–4924.

Dellaert F and Kaess M (2017), "Factor Graphs for Robot Perception," *Foundations and Trends® in Robotics*, **6**(1-2): 1–139.

Eurocode 8 (2004), *Design of Structures for Earthquake Resistance: Part 1: General Rules, Seismic Actions and Rules for Buildings*, European Committee for Standardization.

Forster C, Carlone L, Dellaert F and Scaramuzza D (2016), "On-Manifold Preintegration for Real-Time Visual–Inertial Odometry," *IEEE Transactions on Robotics*, **33**(1): 1–21.

Guan X, Burton H and Shokrabadi M (2021), "A Database of Seismic Designs, Nonlinear Models, and Seismic Responses for Steel Moment-Resisting Frame Buildings," *Earthquake Spectra*, **37**(2): 1199–1222.

Helmberger M, Morin K, Berner B, Kumar N, Cioffi G and Scaramuzza D (2022), "The Hilti Slam Challenge Dataset," *IEEE Robotics and Automation Letters*, **7**(3): 7518–7525.

Hsu TY and Kuo XJ (2020), "A Stand-Alone Smart Camera System for Online Post-Earthquake Building Safety Assessment," *Sensors*, **20**(12): 3374.

Huang Q, Papalia A and Leonard JJ (2022a), "Nested Sampling for Non-Gaussian Inference in SLAM Factor Graphs," *IEEE Robotics and Automation Letters*, **7**(4): 9232–9239.

Huang Q, Pu C, Khosoussi K, Rosen DM, Fourie D, How JP and Leonard JJ (2022b), "Incremental Non-Gaussian Inference for SLAM Using Normalizing Flows," *IEEE Transactions on Robotics.*

Kalaitzakis M, Cain B, Carroll S, Ambrosi A, Whitehead C and Vitzilaios N (2021), "Fiducial Markers for Pose Estimation," *Journal of Intelligent & Robotic Systems*, **101**(4): 1–26.

Kallwies J, Forkel B and Wuensche H-J (2020), "Determining and Improving the Localization Accuracy of AprilTag Detection," *In*: *2020 IEEE International Conference on Robotics and Automation* (*ICRA*), pp. 8288–8294.

Khelifi F, Bradai A, Benslimane A, Rawat P and Atri M (2019), "A Survey of Localization Systems in Internet of Things," *Mobile Networks and Applications*, **24**: 761–785.

Levine NM and Spencer Jr BF (2022), "Post-Earthquake Building Evaluation Using UAVs: A BIM-Based Digital Twin Framework," *Sensors*, **22**(3): 873.

Lutz P, Schuster MJ and Steidle F (2019), "Visual-Inertial SLAM Aided Estimation of Anchor Poses and Sensor Error Model Parameters of UWB Radio Modules," In: *2019 19th International Conference on Advanced Robotics* (*ICAR*), pp. 739–746.

Mao Z, Yan Y, Wu J, Hajjar JF and Padlr T (2018), "Towards Automated Post-Disaster Damage Assessment of Critical Infrastructure with Small Unmanned Aircraft Systems," In*: 2018 IEEE International Symposium on Technologies for Homeland Security* (*HST*), pp. 1–6.

Mazzoni S, McKenna F, Scott MH, Fenves GL, *et al*. (2006), "OpenSees Command Language Manual," *Pacific Earthquake Engineering Research Center* (*PEER*), **264**(1): 137–158.

McCallen D, Petrone F, Coates J and Repanich N (2017), "A Laser-Based Optical Sensor for Broad-Band Measurements of Building Earthquake Drift," *Earthquake Spectra*, **33**(4): 1573–1598.

Munoz-Salinas R and Medina-Carnicer R (2020), "UcoSLAM: Simultaneous Localization and Mapping

by Fusion of Keypoints and Squared Planar Markers," *Pattern Recognition*, **101**: 107193.

Nguyen TH, Nguyen TM and Xie L (2020a), "Tightly-Coupled Single-Anchor Ultra-Wideband-Aided Monocular Visual Odometry System," In: *2020 IEEE International Conference on Robotics and Automation* (*ICRA*), pp. 665–671.

Nguyen TH, Nguyen TM and Xie L (2020b), "Tightly-Coupled Ultra-Wideband-Aided Monocular Visual SLAM with Degenerate Anchor Configurations," *Autonomous Robots*, **44**(8): 1519–1534.

Nguyen TH, Nguyen TM and Xie L (2021a), "Range-Focused Fusion of Camera-IMU-UWB for Accurate and Drift-Reduced Localization," *IEEE Robotics and Automation Letters*, **6**(2): 1678–1685.

Nguyen TM, Cao M, Yuan S, Lyu Y, Nguyen TH and Xie L (2021b), "LIRO: Tightly Coupled Lidar-Inertia-Ranging Odometry," In: *2021 IEEE International Conference on Robotics and Automation* (*ICRA*), pp. 14484–14490.

Nguyen TM, Cao M, Yuan S, Lyu Y, Nguyen TH and Xie L (2021c), "Viral-Fusion: A Visual-Inertial-Ranging-Lidar Sensor Fusion Approach," *IEEE Transactions on Robotics*, **38**(2): 958–977.

Nguyen TM, Yuan S, Cao M, Lyu Y, Nguyen TH and Xie L (2021d), "NTU VIRAL: A Visual-Inertial-Ranging-Lidar Dataset, from an Aerial Vehicle Viewpoint," *The International Journal of Robotics Research*, **41**(3): 270–280.

NZS 1170.5 (2004), *Structural Design Actions Part 5* : *Earthquake Actions - New Zealand*, New Zealand.

Olson E (2011), "AprilTag: A Robust and Flexible Visual Fiducial System," In: *2011 IEEE International Conference on Robotics and Automation*, pp. 3400–3407.

Oppermann I, Hämäläinen M and Iinatti J (2004), *UWB: Theory and Applications*, John Wiley & Sons, USA.

Pacific Earthquake Engineering Research Centre (PEER) (2017), *Tall Buildings Initiative (TBI) - Guidelines for Performance-Based Seismic Design of Tall Buildings*, Berkeley: University of California, Berkeley, CA, USA.

Park SW, Park HS, Kim JH and Adeli H (2015), "3D Displacement Measurement Model for Health Monitoring of Structures Using a Motion Capture System," *Measurement*, **59**: 352–362.

Petrone F, McCallen D, Buckle I and Wu S (2018), "Direct Measurement of Building Transient and Residual Drift Using an Optical Sensor System," *Engineering Structures*, **176**: 115–126.

Priyantha NB, Balakrishnan H, Demaine ED and Teller S (2005), "Mobile-Assisted Localization in Wireless Sensor Networks," In: *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, pp. 172–183.

Ramezani M, Khosoussi K, Catt G, Moghadam P, Williams J, Borges P, Pauling F and Kottege N (2022), "Wildcat: Online Continuous-Time 3D Lidar-Inertial SLAM," *arXiv*, *Preprint arXiv*:220512595.

Recchiuto CT and Sgorbissa A (2018), "Post-Disaster Assessment with Unmanned Aerial Vehicles: A Survey on Practical Implementations and Research Approaches," *Journal of Field Robotics*, **35**(4): 459–490.

Romero-Ramirez FJ, Muñoz-Salinas R and Medina-Carnicer R (2018), "Speeded up Detection of Squared Fiducial Markers," *Image and Vision Computing*, **76**: 38–47.

Shi J, Yang H and Carlone L (2021), "ROBIN: a Graph-Theoretic Approach to Reject Outliers in Robust Estimation Using Invariants," In: *2021 IEEE International Conference on Robotics and Automation* (*ICRA*), pp. 13820–13827.

Skolnik DA and Wallace JW (2010), "Critical Assessment of Interstory Drift Measurements," *Journal of Structural Engineering*, **136**(12): 1574–1584.

Stachniss C, Leonard JJ and Thrun S (2016), "Simultaneous Localization and Mapping," *Springer Handbook of Robotics*, pp. 1153–1176.

Torok M, Fard M and Kochersberger K (2012), "Post-Disaster Robotic Building Assessment: Automated 3D Crack Detection from Image-Based Reconstructions," In: *Computing in Civil Engineerin*g, pp. 397–404.

Torok MM, Golparvar-Fard M and Kochersberger KB (2014), "Image-Based Automated 3D Crack Detection for Post-Disaster Building Assessment," *Journal of Computing in Civil Engineering*, **28**(5): A4014004.

Wang C, Zhang H, Nguyen TM and Xie L (2017), "Ultra-Wideband Aided Fast Localization and Mapping System," In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), pp. 1602–1609.

Wang S, Rodgers C, Zhai G, Matiki TN, Welsh B, Najafi A, Wang J, Narazaki Y, Hoskere V and Spencer Jr BF (2022), "A Graphics-Based Digital Twin Framework for Computer Vision-Based Post-Earthquake Structural Inspection and Evaluation Using Unmanned Aerial Vehicles," *Journal of Infrastructure Intelligence and Resilience*, **1**(1): 100003.

Watanabe K (2018), "ARIADNE with Ambiguity Resolution: Visual Marker Based Rapid Initialization of PPP-AR," In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), pp. 7362–7368.

Wisth D, Camurri M and Fallon M (2023), "VILENS: Visual, Inertial, Lidar, and Leg Odometry for All-Terrain Legged Robots", *IEEE Transactions on Robotics*, **39**(1): 309–326.

Yang H, Antonante P, Tzoumas V and Carlone L (2020a), "Graduated Non-Convexity for Robust Spatial Perception: From Non-Minimal Solvers to Global Outlier Rejection," *IEEE Robotics and Automation Letters*, **5**(2): 1127–1134.

Yang H and Carlone L (2023), "Certifiably Optimal Outlier-Robust Geometric Perception: Semidefinite Relaxations and Scalable Global Optimization,"

IEEE Transactions on Pattern Analysis and Machine Intelligence, **45**(3): 2816–2834.

Yang YS, Xue Q, Chen PY, Weng JH, Li CH, Liu CC, Chen JS and Chen CT (2020b), "Image Analysis Applications for Building Inter-Story Drift Monitoring," *Applied Sciences*, **10**(20): 7304.

Yazgan U and Dazio A (2012), "Post-Earthquake Damage Assessment Using Residual Displacements," *Earthquake Engineering & Structural Dynamics*, **41**(8): 1257–1276.

Yick J, Mukherjee B and Ghosal D (2008), "Wireless Sensor Network Survey," *Computer Networks*, **52**(12): 2292–2330.

Zhang Y and Burton HV (2019), "Pattern Recognition Approach to Assess the Residual Structural Capacity of Damaged Tall Buildings," *Structural Safety*, **78**: 12–22.

Zhang Y, Burton HV, Shokrabadi M and Wallace JW (2019), "Seismic Risk Assessment of a 42-Story Reinforced Concrete Dual-System Building Considering Mainshock and Aftershock Hazard," *Journal of Structural Engineering*, **145**(11): 04019135.

Zhang Y, Burton HV, Sun H and Shokrabadi M (2018), "A Machine Learning Framework for Assessing Post-Earthquake Structural Safety," *Structural Safety*, **72**: 1–16.