Encoder-Camera-Ground Penetrating Radar Sensor Fusion: Bimodal Calibration and Subsurface Mapping

Chieh Chou, Haifeng Li, *Member, IEEE*, and Dezhen Song, *Senior Member, IEEE*

***Abstract*—In this article, we report system and algorithmic developments for a sensing suite comprising a camera and a ground penetrating radar (GPR) with a wheel encoder designed for both surface and subsurface infrastructure inspection, which is a multimodal mapping task. To fuse different sensor modalities properly, we solve a novel GPR-camera calibration problem and a synchronization-challenged sensor fusion problem. First, we design a calibration rig, model the GPR imaging process, introduce a mirror to obtain the joint coverage between the camera and the GPR, and employ the maximum-likelihood estimator to estimate the relative pose between the camera and the GPR with error analysis. Second, we propose a data collection scheme using the customized artificial landmarks to synchronize camera images (temporally evenly spaced) and GPR/encoder data (spatially evenly spaced). We also employ pose graph optimization with location discrepancy as penalty functions to perform data fusion for 3-D reconstruction. We have tested our system in physical experiments. The results show that our system successfully fuses encoder–camera–GPR sensory data and accomplishes a metric 3-D reconstruction. Moreover, our sensor fusion approach reduces the end-to-end distance error from 6.4 to 0.7 cm in a real bridge inspection experiment if comparing to the counterpart that only uses encoder measurements.**

***Index Terms*—Calibration, ground penetrating radar (GPR), mapping, sensor fusion, vision.**

I. Introduction

INFRASTRUCTURES such as bridge decks, freeways, and airport runways require periodic inspections for maintenance purposes due to deterioration over time. Manual inspections would be labor intensive and costly. A more viable approach is to mount sensors onto a robot to perform the inspection tasks. The inspection tasks require both surface and subsurface mapping to assist searching for cracks, voids, or other damages, and thus the ability to combine surface images with subsurface scans is important for further inspections or future repairs (e.g., Fig. 1). Therefore, we combine multiple inspection sensors, such as a regular camera, a light detection and ranging (LIDAR) device, and a ground penetrating radar (GPR), together along with navigational sensors, such as a wheel encoder and/or a global position system (GPS) receiver, into a multimodal sensing suite.

To perform sensor fusion using our sensing suite, we employ the camera not only for surface inspection but also for visual simultaneous localization and mapping (vSLAM), which provides more accurate pose estimation than that from a GPS receiver in a local region. By combining GPR scannings with vSLAM results, GPR applications are no longer restricted to a flat surface due to its inability to obtain pose information from its wheel encoder alone. However, combining data from the heterogeneous sensors is challenging. First, unlike the camera,

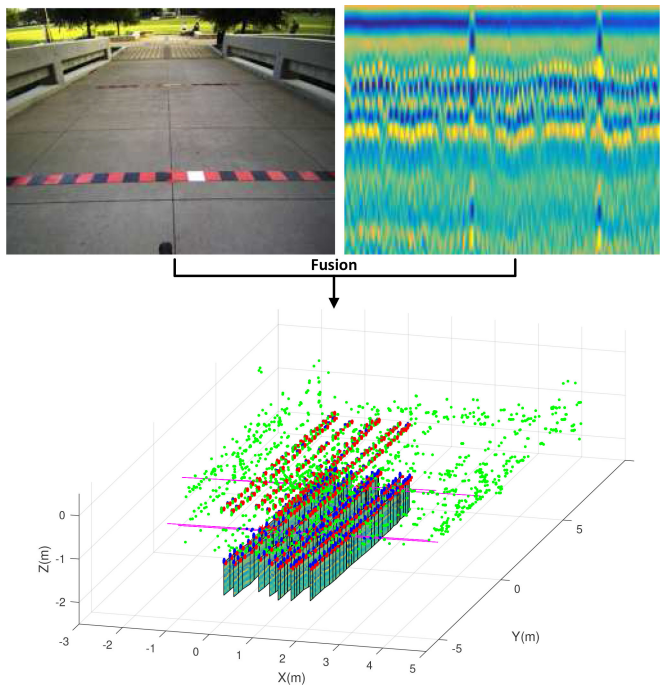


Fig. 1. Inputs: top left: camera image sequences; and top right: a set of GPR scans. Outputs are shown at the bottom: the bridge deck 3-D reconstruction result (including camera/ground penetrating radar (GPR) poses, 3-D landmark positions, artificial landmarks (ALs), and GPR readings) after performing sensor fusion (best viewed in color). Subsurface GPR readings are displayed in the 3-D Euclidean system reconstructed from the camera and the encoder. Detailed 3-D reconstruction results with zoom-in views and explanations are included in the attached video file.

Manuscript received October 24, 2019; revised April 13, 2020; accepted July 7, 2020. Date of publication August 6, 2020; date of current version February 4, 2021. This work was supported in part by the National Science Foundation under Grant NRI-1526200, Grant NRI-1748161, and Grant NRI-1925037. This article was recommended for publication by Associate Editor S. Shen and Editor F. Chaumette upon evaluation of the reviewers' comments. *(Corresponding author: Dezhen Song.)*

Chieh Chou and Dezhen Song are with the Department of Computer Science & Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: jc491986@gmail.com; dzsong@cs.tamu.edu).

Haifeng Li is with Computer Science Department, Civil Aviation University of China, Tianjin 300300, China, and also with the Fujian Provincial Key Laboratory of Informaton Processing and Intelligent Control, Minjiang University, Fuzhou 350108, China (e-mail: hfli@cauc.edu.cn).

This article has supplementary downloadable material available at https://ieeexplore.ieee.org, provided by the authors.

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRO.2020.3010640

the intrinsic 3-D coordinate system of the GPR is not necessarily Euclidean due to its terrain following nature. When the road surface is nonplanar, its output cannot be directly aligned with Euclidean 3-D structure constructed from the camera. This poses a difficult calibration problem. Second, since camera images are taken according to a fixed time interval and GPR scans are triggered by its wheel encoder based on a preset fixed distance traveled, it is nontrivial to synchronize the two sensors.

To enable the heterogeneous sensor suite, we need to solve both the calibration and the synchronization issues. First, we develop a calibration rig and introduce a mirror to allow the joint coverage between the camera and the GPR. We employ a maximum-likelihood estimator (MLE) to obtain the relative pose between the GPR and the camera. Second, we design a data collection scheme using patterned metal plates as ALs to synchronize two types of data streams: the temporally evenly spaced camera images and the spatially evenly spaced GPR/encoder readings. Our algorithm uses ALs to align data streams through pose graph optimization. We have tested our system in physical experiments. The results show that both the calibration and the synchronization designs are successful, which results in the improved mapping accuracy. Comparing to the counterpart that uses encoder measurements only, our algorithm reduces the end-to-end distance error from 6.4 to 0.7 cm in a real bridge inspection experiment.

The remainder of this article is organized as follows. We begin with a review of related work in Section II. The system design is presented in Section III. We introduce the GPR sensing models in Section IV. The camera-GPR calibration is introduced in Section V, followed by bimodal mapping in Section VI. Experiments are in Section VII. Section VIII concludes this article.

II. Related Work

Our proposed system relates to the areas of calibration, GPR applications, vSLAM, and sensor fusion.

Calibration is an important technique to improve the accuracy of a mechanism or a sensor. It usually contains three main components: a model, measurements, and a parameter estimation process [1]. It begins with a closed-form geometry and/or physical model that characterizes the mechanism or the sensing phenomenon. A calibration process is to collect measurements to estimate the model parameters. The measurements are always noisy, which is often described by statistical error models. The noise distribution models can be obtained either analytically or statistically. A Gaussian distribution is a common error model due to its robust asymptotic probability attributes in large populations [2]. The parameter estimation process finds the model parameters by minimizing an aggregated error metric function. Mechanism calibration often solves the kinematic parameters and the inertial parameters for mechanisms with prismatic or revolute joints including robot manipulator calibration [3], pan-tilt robotic cameras calibration [4], and hand–eye calibration [5]. Sensor calibration differs from mechanism calibration due to the unique combination of intrinsic calibration and extrinsic calibration. While the extrinsic model is the similar six degrees of freedom (DoFs) rigid body transformation, the intrinsic model describes the underlying physical principles for the sensing [6] process. Depending on different sensors, the corresponding calibration model varies and leads to different calibration problems, such as camera calibration [7], subsurface pipeline mapping [8], [9], and LIDAR calibration [10].

A GPR is able to measure the time between echoes of electromagnetic signals to perform subsurface survey [11]–[13] with many important applications, such as archeology [14], mine detection and removal [15], [16], bridge deck inspection and evaluation [17]–[20], and planetary exploration [21], [22]. To address the limitation that GPR scanning has to be on a flat surface, we combine the GPR with vSLAM to extend its application range. The vSLAM problem is to simultaneously estimate robot pose and landmark positions using one or more cameras [23], [24]. Without loss of generality, we employ the popular ORB-SLAM2 [25] as a preprocessing step to estimate landmark positions and camera poses

Taking advantage of the complementary nature of different sensory modalities, sensor fusion can improve sensing accuracy, increase robustness, and reduce noises. Related sensor fusion works are camera-LIDAR fusion [26], [27]. Since it is a well-studied area, we skip it in this article. As our focus, camera-GPR sensor fusion has not been well studied yet.

Our group focuses on robotic infrastructure inspection algorithm and system development. This article improves our previous conference papers [28]–[30] by adding pulse response extraction, extending new GPR imaging process regarding patterned metal plates (i.e., ALs), improving accuracy, and conducting more field experiments.

III. System Design

Let us begin with our system design by introducing our sensing suite and the calibration rig.

A. Sensing Suite Design

To enable both surface and subsurface inspection of the infrastructure, we design a sensing suite comprising a camera, a GPR, and a laptop computer as shown in Fig. 2(a). The GPR includes a control unit, a wheel encoder, and an antenna.

1) Configuration: Due to the coverage requirement, the camera needs to be mounted at least 1.0 m above the ground to inspect surface cracks. To scan the subsurface cracks, the GPR antenna needs to be installed close to the ground surface to ensure good radar signal penetration of the concrete structure below. The penetration depth of GPR is inversely proportional to the radio signal frequency, whereas the resolution of GPR scans is proportional to the frequency. Therefore, we choose a 1.6-GHz GPR transceiver antenna because it ensures a 2-m penetration depth for concrete decks with a resolution of less than 4.7 cm. All sensors are mounted on a standard survey cart (Model 623) from Geophysical Survey Systems, Inc.

2) Sensor Choices: The camera is an industry grade 10 mega-pixel CMOS camera (DS-CFMT1000-H); it is applied not only to surface inspection but also to vSLAM because it can provide more accurate trajectory estimation than that from

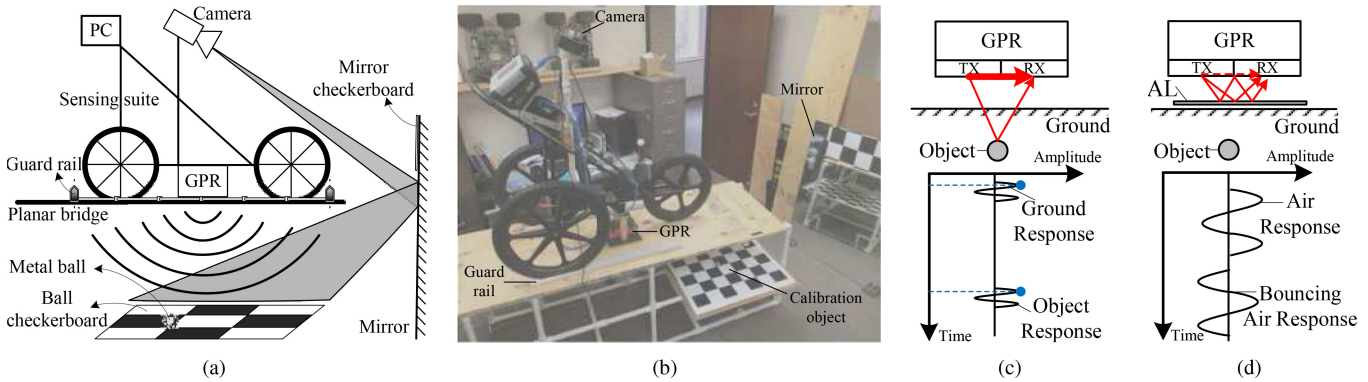


Fig. 2. (a) Camera-GPR sensing suite and calibration schematics. (b) Photo of our sensing suite and calibration rig design. (c) GPR working principles of typical ground and object responses and its A-scan. (d) GPR working principles of metal plate responses and its A-scan.

a GPS receiver in local region. In addition, GPS may not always be available due to terrain or high-rise buildings.

The GPR is used for substructure inspection, and is obtained from GSSI with a control unit SIR-3000. It is worth noting that the wheel encoder from the GPR is also an important sensor for fusing the data. According to [31], the survey wheel has an encoder that sends a fixed number of pulses per revolution to the control unit, and then the control unit uses these pulses to trigger the antenna at equal distance intervals. Therefore, since the GPR readings are triggered by the wheel encoder based on a preset fixed distance traveled, all GPR scans are indexed by wheel encoder pluses. Hence, the GPR data are presynchronized with the wheel encoder data.

B. Calibration Rig Design

We design a calibration rig to cope with challenges brought by the two different modalities and their disjoint coverage [see Fig. 2(b)].

1) Calibration Object Design, Joint Coverage, and Dual Modality Signal Registration: We choose metal balls as the calibration object for the GPR because they are insensitive to orientations and have good reflections to radar signals. The planar black and white wooden checkerboard pattern is chosen to calibrate the camera since it can be easily perceived by the camera but does not interfere with radar signals.

To ensure joint coverage of the two sensors, we install a planar mirror in front of our artificial bridge to create the joint coverage. Also, we attach another checkerboard on the mirror to estimate the mirror pose. As shown in Fig. 2(a) and (b), we name the checkerboard on the mirror as mirror checkerboard and the checkerboard under the artificial bridge as ball checkerboard.

To obtain the relative pose, it is necessary for the two sensors to detect the collocated calibration objects in the joint coverage space. For the collocated calibration objects, we place the metal ball on top of the ball checkerboard pattern because the camera and the GPR can detect them respectively. The metal ball center is always directly above the a checkerboard vertex with its radius as the distance between the metal ball center and the ball checkerboard.

2) Ensure Euclidean Property for the GPR: The challenge arises from the fact that a GPR is a terrain following sensor, and the geometric relationship in the Euclidean space can only be easily interpreted from a GPR image if the GPR travels along a straight line on a plane. If we calibrate the GPR traveling on an arbitrary surface, we cannot obtain proper coordinate system transformation from a non-Euclidean system to the camera coordinate system. To ensure Euclidean sensory data, a GPR must move on a planar surface along a straight linear trajectory during calibration. Also, our calibration rig is designed to be a planar artificial bridge with two straight guard rails (one for each rear wheel) to ensure that the GPR cart follows straight line motion during the scanning process. Furthermore, our guard rails are equipped with mechanical stops to ensure the repeatability for both sensors to collect data at each stop.

IV. GPR Sensing Models

Now, let us explain the working principles of a GPR with response extraction, and then model the GPR imaging process for a metal ball (i.e., calibration object) and a metal plate (i.e., AL).

A. Principle of Operations

Our GPR has a bistatic antenna that contains a transmitter (TX) and a receiver (RX) for radar signals [31], [32]. Due to the antenna's close proximity to ground, our GPR can be described by using a ground-coupled model [33], [34] [see Fig. 2(c)]. During the GPR operation, TX emits pulses to the subsurface and RX detects the reflected pulses if the emitted pulses encounter objects or layers with different dielectric permittivity. Then, the GPR estimates the traveling time between the emitted pulses and the echoed pulses, and generate an A-scan that records the signal amplitude versus traveling time at this GPR position. Fig. 2(c) bottom coordinates illustrate an example of A-scan. The A-scan usually has two main types of responses: the ground response and the object response. The ground response denotes a direct transmission from TX to RX [see the thick red arrow in Fig. 2(c)]. The object response indicates that the signal transmits from TX to the object and then bounces back to RX [see the thin red arrow in Fig. 2(c)]. By extracting the ground and the object responses

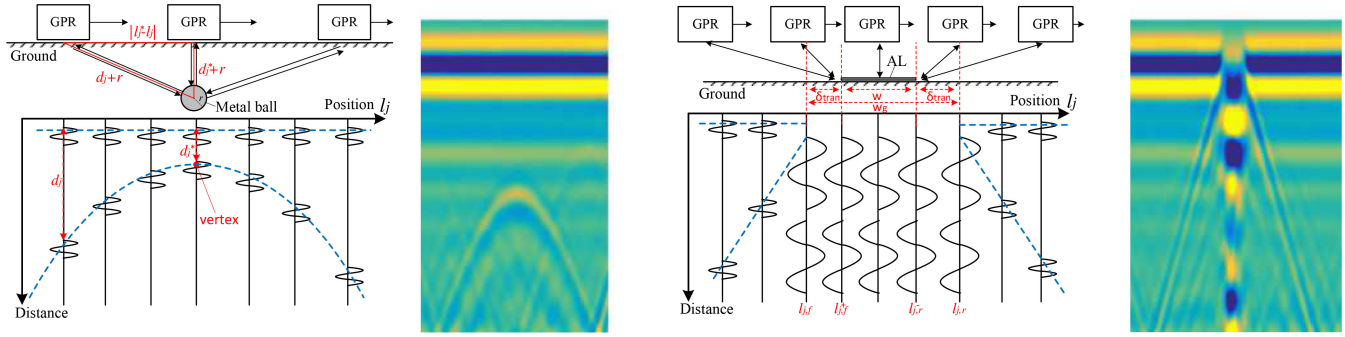


Fig. 3. (a) Illustration of the hyperbola response regarding a metal ball (calibration object). (b) Real GPR image of (a). (c) Illustration of the rectangle response regarding a metal plate. (d) Real GPR image of (c).

in the A-scan, we can estimate the distance from the GPR to the object.

B. Assumptions and Notations

To model the GPR sensing modality and the GPR imaging process, we assume that the following statements hold.

1) The dielectric material is uniformly dielectric and the radio wave propagation velocity is a known constant v_c in the homogeneous medium.
2) The TX-RX offset is a known intrinsic parameter and pre-compensated in GPR scans. When estimated the distance from the GPR to the object, the TX–RX offset is negligible if comparing to the GPR–object distance.

The GPR scans while moving, and it collects the A-scans at each point defined by preset wheel encoder intervals on the trajectory to assemble a B-scan as shown in Fig. 3(a). Let us define common notations as follows.

1) $A_j = \{a_{t,j} | t = 1, 2, \ldots, n_r\}$ denotes the jth A-scan, where t is the traveling time, $a_{t,j}$ is the signal amplitude, and n_r is the number of readings.
2) l_j denotes the GPR's traveled distance measured by the wheel encoder from the first A-scan to the jth A-scan.
3) $B = \{\{A_j, l_j\} | j = 1, \ldots, m\}$ denotes a B-scan where m is the total number of A-scans. Since this definition is very similar to an image coordinate system, we also call B as GPR image in this article.

C. Pulse Response Extraction

Based on the assumptions in Section IV-B, the traveling times between the first peaks of ground response and object response [see the blue points in Fig. 2(c)] can be approximately estimated as the two times distance between the GPR and the object according to [11], [31]. To obtain peaks of the two types of response, we need to extract the pulse responses in each $A_j \in B$. To properly extract the peak positions, we apply damped sinusoidal model [14], [35] given by

$$a_{t,j} = m(\alpha_j, \beta_j, \gamma_j, \omega_j, \phi_j) = \beta_j e^{-\alpha_j t} \cos(\omega_j t + \phi_j) + \gamma_j \tag{1}$$

where $a_{t,j}$ is the signal amplitude and m is the damped sinusoidal model with parameters defined as follows: β_j is the amplitude, ω_j is the angular frequency, ϕ_j is the phase, γ_j is the offset for the model, and α_j is the attenuation constant [36] defined by

$$\alpha_j = \omega_j \sqrt{\frac{\mu_j \epsilon_j}{2}} \left[\left(1 + \left(\frac{\sigma_j}{\omega_j \epsilon_j} \right)^2 \right)^{1/2} - 1 \right]^{1/2} \approx \frac{\sigma_j}{2} \sqrt{\frac{\mu_j}{\epsilon_j}} \tag{2}$$

where μ_j is the magnetic permeability, ϵ_j is the dielectric permittivity, and σ_j is the electrical conductivity. The approximation is true due to $\sigma_j \ll \omega_j \epsilon_j$ in our applications, and $\{\mu_j, \epsilon_j, \sigma_j\}$ are the parameters defined by material properties. Of course, we might not know the exact value of α_j due to mixture of materials but it is a constant and can also be estimated in the following.

To estimate the pulses for ground and object in A_j, we use $\{(t, a_{t,j}) | t \in \mathcal{P}\}$ as the measurements, where $\mathcal{P}$ is the index set containing all measurements in the pulse, to formulate an optimization problem

$$\min_{\hat{\alpha}_j, \hat{\beta}_j, \hat{\gamma}_j, \hat{\omega}_j, \hat{\phi}_j} \sum_{t \in \mathcal{P}} ||m(\hat{\alpha}_j, \hat{\beta}_j, \hat{\gamma}_j, \hat{\omega}_j, \hat{\phi}_j) - a_{t,j}||_{\Sigma}^2 \tag{3}$$

where the hat $\hat{}$ indicates the estimators and $\|\cdot\|_{\Sigma}$ denotes Mahalanobis distance.

After obtaining the parameters, we are able to estimate the peak positions of each pulse and obtain the corresponding traveling times. Let $t_{g,j}$ and $t_{o,j}$ be the traveling time corresponding to the peak positions for the ground response and the object response, respectively. GPR data processing converts $t_{g,j}$ and $t_{o,j}$ into the traveling distance $d_{g,j}$ and $d_{o,j}$ by applying intrinsic model using the dielectric permittivity.

D. Hyperbola Response Regarding a Metal Ball

When the GPR senses a metal ball with radius r as shown in Fig. 3(a), the GPR image shows a line (i.e., ground responses) and a hyperbola (e.g. ball responses) as shown in Fig. 3(a) (see the dotted blue curves) and Fig. 3(b). We define $\mathbf{x}_j = [l_j, d_j]^{\mathsf{T}} \in A_j$ as a point in the hyperbola, where $d_j = |d_{o,j} - d_{g,j}|$ is the distance between the ground response and the object response in A_j as shown in Fig. 3(a) bottom. Let $\mathbf{H} = \{\mathbf{x}_j | j = 1, 2, \ldots, m\}$ denotes all the points on the hyperbola corresponding to the metal ball. Let $\mathbf{x}_{j^*} = [l_{j^*}, d_{j^*}]^{\mathsf{T}}$ be the hyperbola vertex, where

j^* indicates the A-scan index corresponding to the hyperbola vertex. When the GPR has moved l_{j^*} and is generating A_{j^*}, d_{j^*} is the shortest distance from the GPR to the metal ball.

When the GPR moves from A_j to A_{j^*}, we are based on the red right triangle shown in Fig. 3(a) top and obtain

$$(d_j + r)^2 = (d_{j^*} + r)^2 + (l_j - l_{j^*})^2. \tag{4}$$

Let $\tilde{d_{j^*}} = d_{j^*} + r$ and $\tilde{d_j} = d_j + r$, and then we sort out (4) in a hyperbola form as

$$\tilde{\mathbf{x}}_j^{\mathsf{T}} \mathbf{Q} \tilde{\mathbf{x}}_j = 0 \quad \forall \mathbf{x}_j \in \mathbf{H} \tag{5}$$

where $\tilde{\mathbf{x}}_j = [l_j, \tilde{d_j}, 1]^{\mathsf{T}}$ and $\mathbf{Q} = \begin{bmatrix} 1 & 0 & -l_{j^*} \\ 0 & -1 & 0 \\ -l_{j^*} & 0 & l_{j^*}^2 + \tilde{d_{j^*}}^2 \end{bmatrix}$.

For a general conic equation $ax^2 + bxy + cy^2 + dx + ey + f = 0$ with five DoFs, the DoFs of (5) in our case decreases to two because $b = e = 0$ and $a = -c = 1$. Therefore, given r, two parameters $\{l_{j^*}, d_{j^*}\}$ in (5) are sufficient to define the corresponding hyperbola. We will detail it in Section V-B.

E. Rectangle Response Generated by a Metal Plate

When the GPR is traveling over a metal plate (a perfect electric conductor with very good reflection for radar signals) on the ground, the ground-coupled model is no longer applicable because the metal plate blocks the coupling and reflects all the transmitting energy without any signal penetration as shown in Fig. 2(d). The reflected signal energy not only generates a lot of signal paths [see the thin red arrows in Fig. 2(d)] because of reflected signals bouncing between the antenna and the plate, but also dominates and shields the direct signal from TX to RX [see the dotted red arrow in Fig. 2(d)] due to its higher signal strength. This leads to the repeated pattern in A-scans as shown in Fig. 2(d) bottom and Fig. 3(d). We identify the shortest traveling time from TX plate to RX as the air response, and others as bouncing air responses. Since all traveling times are longer than that of the direct coupling, this explains why AL responses are slightly slower in the A-scans.

To properly model the GPR signal response when traveling over a metal plate, we have to consider the shielding effect. The shielding effect generates an offset when measuring the width of the metal plate in the GPR image because the shielding does not exactly occur on the margin of the plate as shown in Fig. 3(c). Therefore, we have to compensate the offset when traversing the metal plates. Let δ_{tran} be the offset for both front and rear directions. Let $w_g = |l_{j,r} - l_{j,f}|$ be the width of the metal plate measured by the wheel encoder in GPR, where $l_{j,f}$ and $l_{j,r}$ are the accumulated distances in the first and the last detected A-scans during the shielding. Let w be the width of the metal plate measured by a ruler. Here, we treat w as a constant because its measurement error is negligible comparing to other variables. Assuming the ground surface, the metal plate, and the GPR trajectory are parallel during the shielding, we model the offsets by $w_g = w + 2\delta_{\text{tran}}$. To estimate the optimal solution for δ_{tran}, we formulate an optimization problem

$$\min_{\hat{\delta}_{\text{tran}}} \sum^{n_{\text{AL}}} ||w + 2\hat{\delta}_{\text{tran}} - w_g||_{\Sigma}^2 \tag{6}$$

where $\|\cdot\|_{\Sigma}$ denotes Mahalanobis distance and n_{AL} denotes the number of ALs. Once we know $\hat{\delta}_{\text{tran}}$, we can obtain the corrected $l_{j^+,f}$ and $l_{j^-,r}$ by

$$l_{j^+,f} = l_{j,f} + \hat{\delta}_{\text{tran}}, \qquad l_{j^-,r} = l_{j,r} - \delta_{\text{tran}} \tag{7}$$

which will be used in Section VI-B.

V. Camera-GPR Calibration

With the understanding of GPR sensing models, we design camera-GPR calibration with checkerboards and a metal ball. We begin with data collection description for n trials. Each trial contains N camera images and one GPR B-scan. In each trial, we first place the metal ball on a vertex corner of the ball checkerboard, and then push the sensing suite to scan the calibration object combo. A complete trial allows the GPR to generate a hyperbola response in a GPR image [see Fig. 3(b)], and the camera takes an image at each mechanical stop, which yields N images. Hence, we adjust the metal ball on n different positions on the ball checkerboard and repeat the trial for each metal ball position. We assume that the following statements hold.

1) Camera and GPR intrinsic parameters are precalibrated.
2) The measurement noises follow Gaussian distribution with zero means.

Some common notations are defined as follows.

1) i denotes the trial index which is also the metal position index.
2) $\{W\}$ and $\{M\}$ denote the 3-D world and the mirror coordinate system, respectively. We interchangeably use "frame" and "coordinate system" throughout this article and let default 3-D frames be right-handed coordinates.
3) ${}^W\mathbf{B}_i \in \mathbb{R}^3$ denotes the coordinates of the ith metal ball center position w.r.t. $\{W\}$, and ${}^W\mathbf{X}_i \in \mathbb{R}^3$ denotes its corresponding point in the checkerboard. Without loss of generality, let the checkerboard plane be $Z = 0$ in the world frame, thus we derive ${}^W\mathbf{B}_i = {}^W\mathbf{X}_i + [0, 0, r]^{\mathsf{T}}$, where r is the metal ball radius. ${}^M\mathbf{X}_p \in \mathbb{R}^3$ denotes the coordinates of the pth point on mirror checkerboard w.r.t. $\{M\}$. As a convention, we will use the left superscript to indicate the reference frame in this article.
4) $\{G_k\}$ denotes the GPR frame at stop k, where its origin is at the GPR antenna center, its *Y*-axis is parallel to the GPR moving direction, its *Z*-axis is perpendicular to the surface plane pointing up, and its *X*-axis is perpendicular to the GPR moving direction.
5) $\{C_k\}$ denotes the camera frame at stop k, where its origin is at the camera optical center, its *Z*-axis is coinciding with the optical axis and pointing to the forward direction of the camera, and its *X*-axis and *Y*-axis are parallel to the horizontal and vertical directions of the charge-coupled device (CCD) sensor plane, respectively.

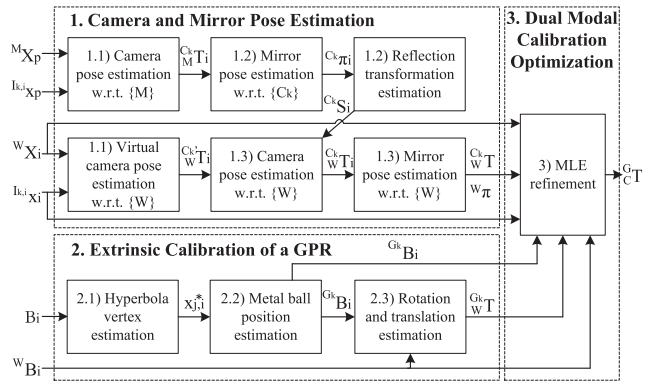


Fig. 4. Calibration computation flow diagram.

6) B_i denotes the GPR image, which is generated by GPR scanning the ith metal ball. In each B_i, we collect N camera images at N stops, and denote $I_{k,i}$ as the kth camera image in the trial.
7) ${}^{G_k}_{C_k}\mathbf{T}$ denotes the rigid body transformation from $\{C_k\}$ to $\{G_k\}$. Since both sensors are fixed firmly on the sensing suite, ${}^{G_1}_{C_1}\mathbf{T}, .., {}^{G_k}_{C_k}\mathbf{T}, ..., {}^{G_N}_{C_N}\mathbf{T}$ are identical. Thus, we denote ${}^{G}_{C}\mathbf{T} = {}^{G_k}_{C_k}\mathbf{T}, \forall k$. As a convention, we use left subscript and superscript to indicate frames in the transformation mapping, and left superscript is the final reference frame.

Now, we define our calibration problem as follows.

Definition 1: Given ${}^{W}\mathbf{B}_i$ with the corresponding B_i and $I_{k,i}$, and given ${}^{M}\mathbf{X}_p$ with their corresponding feature points in $I_{k,i}$, where $i = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, N$, determine ${}^{G}_{C}\mathbf{T}$.

Fig. 4 summarizes our calibration computation flow diagram in three steps. We first compute the initial solution for camera/mirror pose estimation and GPR extrinsic calibration in Sections V-A and V-B, respectively. Finally, we determine the optimal solution by using an optimization approach in Section V-C. We will detail them in the following sections.

A. Camera and Mirror Pose Estimation

First, we want to find the closed-form solution for the frame mapping from world frame to each camera frame ${}^{C_k}_{W}\mathbf{T}$ and mirror plane ${}^{W}\pi$.

1) Camera Pose Estimation from 2-D/3-D Points: Before estimating camera pose, we introduce the camera model. For the ball checkerboard point ${}^{W}\mathbf{X}_i$, the corresponding 2-D image point ${}^{I_{k,i}}\mathbf{x}_i$ in $I_{k,i}$ through the planar mirror can be described by the camera projection model as

$$\begin{bmatrix} {}^{I_{k,i}}\mathbf{x}_i \\ 1 \end{bmatrix} = \lambda \mathbf{K} \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3\times 1} \end{bmatrix} {}^{C'_k}_{W}\mathbf{T} \begin{bmatrix} {}^{W}\mathbf{X}_i \\ 1 \end{bmatrix} \tag{8}$$

where λ is a scalar, $\mathbf{K}$ is the intrinsic camera matrix, and ${}^{C'_k}_{W}\mathbf{T}$ represents the mapping from $\{W\}$ to virtual camera frame $\{C'_k\}$ where the apostrophe $'$ indicates the virtual frame in the mirror. It is worth mentioning that the frames are right-handed coordinates in the real world, but are left handed in the mirrored space. Similarly, for the mirror checkerboard point ${}^{M}\mathbf{X}_p$, the corresponding 2-D image point ${}^{I_{k,i}}\mathbf{x}_p$ in $I_{k,i}$ can be also described by (8) with index $i \to p$, $\{W\} \to \{M\}$, and $\{C'_k\} \to \{C_k\}$.

Based on the camera model and the 2-D/3-D corresponding points, we estimate the rigid body transformation between the reference frame and the camera frame by solving the perspective-n-point (PnP) problem [37], [38]. Therefore, given ${}^{M}\mathbf{X}_p$ and ${}^{I_{k,i}}\mathbf{x}_p$, we estimate ${}^{C_k}_{M}\mathbf{T}$; given ${}^{W}\mathbf{X}_i$ and ${}^{I_{k,i}}\mathbf{x}_i$, we estimate ${}^{C'_k}_{W}\mathbf{T}$.

2) Mirror Pose and Reflection Transformation Estimation W.R.T. $\{C_k\}$: Let ${}^{C_k}\pi = [{}^{C_k}\mathbf{n}^{\mathsf{T}}, {}^{C_k}c]^{\mathsf{T}}$ be the parameters for the planar mirror in $\{C_k\}$, where ${}^{C_k}\mathbf{n}$ is the normal vector of the mirror surface and ${}^{C_k}c$ is the orthogonal distance from the camera origin to the mirror surface. Similarly, ${}^{W}\pi = [{}^{W}\mathbf{n}^{\mathsf{T}}, {}^{W}c]^{\mathsf{T}}$ represents the planar mirror in $\{W\}$.

After receiving ${}^{C_k}_{M}\mathbf{T}$ from Section V-A1, the mirror parameters ${}^{C_k}\pi$ can be obtained by the plane equation ${}^{C_k}\mathbf{n}^{\mathsf{T}}{}^{C_k}\mathbf{X} = {}^{C_k}c$, where ${}^{C_k}\mathbf{X}$ denotes any point on the mirror plane. Let ${}^{C_k}_{M}\mathbf{T} = \begin{bmatrix} {}^{C_k}M\mathbf{R} & {}^{C_k}_{M}\mathbf{t} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}$ and ${}^{M}\mathbf{X}_p$ be on the plane $Z = 0$ in mirror frame $\{M\}$, then ${}^{C_k}\mathbf{n}$ equals to the third column of the rotation matrix ${}^{C_k}_{M}\mathbf{R}$ and ${}^{C_k}c = -{}^{C_k}\mathbf{n}^{\mathsf{T}}({}^{C_k}_{M}\mathbf{R}^{\mathsf{T}}{}^{C_k}_{M}\mathbf{t})$. Finally, according to the planar mirror model [39]–[43], the mirror reflection transformation in $\{C_k\}$ is obtained by

$${}^{C_k}\mathbf{S} = \begin{bmatrix} \mathbf{I}_3 - 2{}^{C_k}\mathbf{n}{}^{C_k}\mathbf{n}^{\mathsf{T}} & 2{}^{C_k}c{}^{C_k}\mathbf{n} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}. \tag{9}$$

3) Camera and Mirror Pose Estimation W.R.T. $\{W\}$: In this section, we want to solve ${}^{C_k}_{W}\mathbf{T}$ and ${}^{W}\pi$. First, we compute ${}^{C_k}_{W}\mathbf{T}$ by using the camera projection model with mirror reflection transformation. Based on ${}^{C'_k}_{W}\mathbf{T}$ in (8) and ${}^{C_k}\mathbf{S}$ in (9), we are able to obtain ${}^{C_k}_{W}\mathbf{T}$ by

$${}^{C_k}_{W}\mathbf{T} = {}^{C_k}\mathbf{S}\,{}^{C'_k}_{W}\mathbf{T}. \tag{10}$$

To increase the accuracy, we average ${}^{C_k}_{W}\mathbf{T}$ over all metal ball center positions i, where $i = 1, 2, \ldots, n$, as follows.

Let ${}^{C_k}_{W}\mathbf{T} = \begin{bmatrix} {}^{C_k}_{W}\mathbf{R} & {}^{C_k}_{W}\mathbf{t} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}$ be the final transformation, we want to estimate and let ${}^{C_k}_{W}\mathbf{T}_i = \begin{bmatrix} {}^{C_k}_{W}\mathbf{R}_i & {}^{C_k}_{W}\mathbf{t}_i \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}$ be the measurements w.r.t. the ith metal ball. We use rotation averaging [44] to find the optimal rotation matrix ${}^{C_k}_{W}\mathbf{R}$ by minimizing the following cost function:

$$\begin{aligned} C({}^{C_k}_{W}\mathbf{R}) &= \sum_{i=1}^{n} \|{}^{C_k}_{W}\mathbf{R}_i - {}^{C_k}_{W}\mathbf{R}\|_F^2 \\ &= C_1 - 2\left\langle ; \sum_{i=1}^{n} {}^{C_k}_{W}\mathbf{R}_i, {}^{C_k}_{W}\mathbf{R} \right\rangle \end{aligned} \tag{11}$$

where $\|\cdot\|_F$ is the Frobenius matrix norm, $\langle \cdot, \cdot \rangle$ is the Frobenius inner product, and C_1 is a constant. Thus, to minimize (11)

becomes

$$\max_{{}^{C_k}_{W}\mathbf{R}\in SO(3)} \left\langle \sum_{i=1}^{n} {}^{C_k}_{W}\mathbf{R}_i, {}^{C_k}_{W}\mathbf{R} \right\rangle = \min_{{}^{C_k}_{W}\mathbf{R}\in SO(3)} \left\| \sum_{i=1}^{n} {}^{C_k}_{W}\mathbf{R}_i - {}^{C_k}_{W}\mathbf{R} \right\|_F^2. \quad (12)$$

As a result, the closed-form solution for ${}^{C_k}_{W}\mathbf{R}$ is given by ${}^{C_k}_{W}\mathbf{R} = \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$, where $\mathbf{U}$ and $\mathbf{V}$ are obtained from singular value decomposition of $\sum_{i=1}^{n} {}^{C_k}_{W}\mathbf{R}_i = \mathbf{U}\Sigma\mathbf{V}^\mathsf{T}$, and $\mathbf{S} = \text{diag}(1, 1, -1)$ if $\det(\mathbf{U}\mathbf{V}^\mathsf{T}) \geq 0$, otherwise $\mathbf{S} = \mathbf{I}_{3\times 3}$ because ${}^{C_k}_{W}\mathbf{R}$ is an improper rotation ($\det({}^{C_k}_{W}\mathbf{R}) = -1$). Also, we estimate the average translation vector by ${}^{C_k}_{W}\mathbf{t} = \sum_{i=1}^{n} {}^{C_k}_{W}\mathbf{t}_i/n$.

Next, we show how to derive ${}^{W}\pi$. Similarly to (9), the mirror reflection transformation in $\{W\}$ is given by

$${}^{W}\mathbf{S} = {}^{C_k}_{W}\mathbf{T}^{-1}{}^{C_k}\mathbf{S}{}^{C_k}_{W}\mathbf{T} = \begin{bmatrix} \mathbf{I}_3 - 2{}^{W}\mathbf{n}{}^{W}\mathbf{n}^\mathsf{T} & 2{}^{W}c{}^{W}\mathbf{n} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}. \quad (13)$$

Once ${}^{W}\mathbf{S}$ is known, ${}^{W}\pi$ can be obtained by decomposing ${}^{W}\mathbf{S}$. Also, to improve the accuracy, we average ${}^{W}\mathbf{S}$ over all metal ball center positions i and all stops k, where $i = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, N$, as follows.

Let ${}^{W}\mathbf{S} = \begin{bmatrix} {}^{W}\mathbf{R} & {}^{W}\mathbf{t} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}$ be the mirror reflection transformation we want to estimate and let ${}^{W}\mathbf{S}_{k,i} = {}^{C_k}_{W}\mathbf{T}_i{}^{-1}{}^{C_k}\mathbf{S}_i{}^{C_k}_{W}\mathbf{T}_i = \begin{bmatrix} {}^{W}\mathbf{R}_{k,i} & {}^{W}\mathbf{t}_{k,i} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}$ be the mirror reflection transformation w.r.t. the ith metal ball at stop k. We decompose ${}^{W}\mathbf{t}_{k,i}$ into ${}^{W}c_{k,i}$ and ${}^{W}\mathbf{n}_{k,i}$ by ${}^{W}\mathbf{n}_{k,i} = {}^{W}\mathbf{t}_{k,i}/\|{}^{W}\mathbf{t}_{k,i}\|$ and ${}^{W}c_{k,i} = \|{}^{W}\mathbf{t}_{k,i}\|/2$. Finally, we obtain ${}^{W}\pi = [{}^{W}\mathbf{n}^\mathsf{T}, {}^{W}c]^\mathsf{T}$ by

$${}^{W}\mathbf{n} = \sum_{i=1}^{n}\sum_{k=1}^{N} \frac{{}^{W}\mathbf{n}_{k,i}}{\|\sum_{i=1}^{n}\sum_{k=1}^{N} {}^{W}\mathbf{n}_{k,i}\|}, \quad {}^{W}c = \sum_{i=1}^{n}\sum_{k=1}^{N} \frac{{}^{W}c_{k,i}}{nN}. \quad (14)$$

B. Extrinsic Calibration of a GPR

Now, let us explain the GPR calibration in Fig. 4 bottom left. We want to find metal ball center positions ${}^{G_k}\mathbf{B}_i$ w.r.t. $\{G_k\}$ and the frame mapping from $\{W\}$ to each GPR frame ${}^{G_k}_{W}\mathbf{T}$.

1) Hyperbola Vertex Estimation: Recall that we defined $\{\mathbf{x}_j, \mathbf{H}, \mathbf{x}_{j^*}\}$ in Section IV-D. We now use subindex i to indicate measurements in B_i, and thus we redefine $\mathbf{x}_{j,i} = [l_{j,i}, d_{j,i}]^\mathsf{T}$ as one point in the hyperbola in B_i, $\mathbf{H}_i$ as all the points on the hyperbola in B_i, and $\mathbf{x}_{j^*,i} = [l_{j^*,i}, d_{j^*,i}]^\mathsf{T}$ as the hyperbola vertex.

To apply MLE to solve $\mathbf{x}_{j^*,i}$, we first model $\mathbf{x}_{j,i}$'s measurement error as a zero mean Gaussian with covariance matrix $\sigma_{j,i}^2\mathbf{I}$. Because r is a constant, the noise distribution of $\tilde{d}_{j,i}$ is the same as that of $d_{j,i}$. By stacking all measurements $\mathbf{x}_{j,i}$ together, we estimate the hyperbola vertex $\mathbf{x}_{j^*,i}$ using the overall error function

$$\phi(\mathbf{x}_{j^*,i}) = \begin{bmatrix} \tilde{\mathbf{x}}_{1,i}^\mathsf{T}\mathbf{Q}_i\tilde{\mathbf{x}}_{1,i} \\ \vdots \\ \tilde{\mathbf{x}}_{m,i}^\mathsf{T}\mathbf{Q}_i\tilde{\mathbf{x}}_{m,i} \end{bmatrix}. \quad (15)$$

Then, the MLE of $\mathbf{x}_{j^*,i}$ is obtained by minimizing

$$\min_{\mathbf{x}_{j^*,i}} \phi(\mathbf{x}_{j^*,i})^\mathsf{T}\Sigma_i^{-1}\phi(\mathbf{x}_{j^*,i}) \quad (16)$$

where $\Sigma_i = \text{diag}(\sigma_{1,i}^2, \sigma_{2,i}^2, \ldots, \sigma_{m,i}^2)$. This nonlinear optimization problem can be solved by the Levenberg–Marquardt (LM) algorithm [45].

2) Metal Ball Center Position Estimation: Let h_i be the vertical distance from the bridge surface to the ith metal ball surface position, and its noise distribution be a zero mean Gaussian with variance $\sigma_{h_i}^2$. Considering the metal ball radius r, we use $\tilde{h}_i = h_i + r$ as the distance from the bridge surface to the ith metal ball center position. The noise distribution of $\tilde{h}_i$ is the same as that of h_i because r is a constant. Let l_k be the GPR traveled distance from the first scan to $\{G_k\}$ where index k indicates the scan when GPR pauses at mechanical stop k. Given $\{\mathbf{x}_{j^*,i}, l_k, \tilde{h}_i\}$, we compute the metal ball coordinates ${}^{G_k}\mathbf{B}_i$ as shown in Fig. 5 with its covariance matrix ${}^{G_k}\Sigma_i$ by

$${}^{G_k}\mathbf{B}_i = \begin{bmatrix} \sqrt{\tilde{d}_{j^*,i}^2 - \tilde{h}_i^2} \\ l_{j^*} - l_k \\ -\tilde{h}_i \end{bmatrix}, {}^{G_k}\Sigma_i = J_G \begin{bmatrix} {}^{D_{k,i}}\Sigma & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & \sigma_{\tilde{h}_i}^2 \end{bmatrix} J_G^\mathsf{T} \quad (17)$$

where $J_G = \frac{\partial {}^{G_k}\mathbf{B}_i}{\partial(\mathbf{x}_{j^*,i}, l_k, \tilde{h}_i)}$ is a Jacobian matrix, ${}^{D_{k,i}}\Sigma = (J_\phi^T\Sigma_i^{-1}J_\phi)^{-1}$ is the covariance matrix of $\{\mathbf{x}_{j^*,i}, l_k\}$, and $J_\phi = \frac{\partial\phi}{\partial(\mathbf{x}_{j^*,i}, l_k)}$ is also a Jacobian matrix.

3) Rotation and Translation Estimation: After obtaining the metal ball coordinates of $\{G_k\}$ and $\{W\}$, we compute the closed-form solution of ${}^{G_k}_{W}\mathbf{T}$ by Horn's method [46] and then refine the solution by MLE. For the error analysis and more details, please refer to [28].

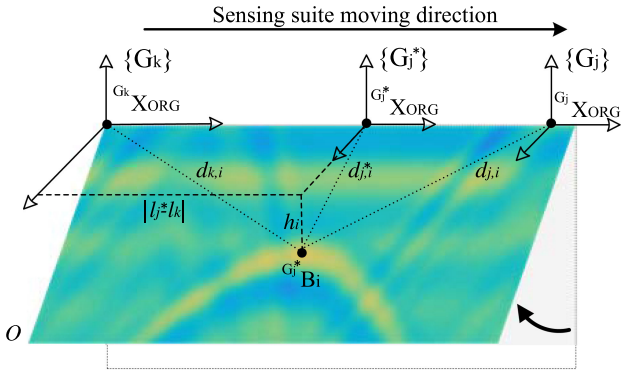


Fig. 5. Illustration of the GPR frames on the surface sensing the ith metal ball. The index j^* indicates the scan when GPR is closest to the object, which also means the hyperbola vertex.

C. Bimodal Calibration Optimization

To achieve the optimal solution by the initial calibration results from Sections V-A and V-B, we formulate an overall optimization problem to estimate ${}^{G}_{C}\mathbf{T}$, ${}^{W}\pi$, $\{{}^{C_k}_{W}\mathbf{T}|k=1,\ldots,N\}$, $\{{}^{W}\mathbf{X}_i|i=1,\ldots,n\}$, and $\{{}^{W}\mathbf{B}_i|i=1,\ldots,n\}$ based on all the measurements. The relationship of the coordinate systems is given by

$$\begin{bmatrix} {}^{G_k}\mathbf{X}_i \\ 1 \end{bmatrix} = {}^{G_k}_{W}\mathbf{T} \begin{bmatrix} {}^{W}\mathbf{X}_i \\ 1 \end{bmatrix} = {}^{G}_{C}\mathbf{T}\,{}^{C_k}_{W}\mathbf{T} \begin{bmatrix} {}^{W}\mathbf{X}_i \\ 1 \end{bmatrix}. \tag{18}$$

Let ${}^{C_k}_{W}\mathbf{T} = \begin{bmatrix} {}^{C_k}_{W}\mathbf{R} & {}^{C_k}_{W}\mathbf{t} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}$ and let the six-vector representation of ${}^{C_k}_{W}\mathbf{R}$ and ${}^{C_k}_{W}\mathbf{t}$ be $\xi_k = [\theta_{x,k}, \theta_{y,k}, \theta_{z,k}, t_{x,k}, t_{y,k}, t_{z,k}]^{\mathsf{T}}$, where $(\theta_{x,k}, \theta_{y,k}, \theta_{z,k})$ is the Euler angle representation of ${}^{C_k}_{W}\mathbf{R}$ in the order of *Z–Y–X*, and ${}^{C_k}_{W}\mathbf{t} = [t_{x,k}, t_{y,k}, t_{z,k}]^{\mathsf{T}}$. The initial value of ${}^{G}_{C}\mathbf{T}$ can be obtained by ${}^{G}_{C}\mathbf{T} = {}^{G_k}_{W}\mathbf{T}\,{}^{C_k}_{W}\mathbf{T}^{-1}$. Let η be the six-vector representation of ${}^{G}_{C}\mathbf{T}$, where the notation definitions are similar to ξ_k. Let ${}^{W}\hat{\mathbf{X}}_i$ and ${}^{W}\hat{\mathbf{B}}_i$ be the estimators of ${}^{W}\mathbf{X}_i$ and ${}^{W}\mathbf{B}_i$, respectively. We define the parameterized function for camera as ${}^{I_{k,i}}\mathbf{x}_i = f({}^{W}\pi, \xi_k, {}^{W}\mathbf{X}_i)$ according to (8) and (13), and for GPR as ${}^{G_k}\mathbf{B}_i = g(\eta, \xi_k, {}^{W}\mathbf{B}_i)$ according to (18).

Finally, let $\mathbf{p} = [\eta^{\mathsf{T}}, \xi^{\mathsf{T}}, {}^{W}\pi^{\mathsf{T}}, \mathbf{X}_c^{\mathsf{T}}, \mathbf{X}_g^{\mathsf{T}}]^{\mathsf{T}}$ be the estimated vector, where $\xi = [\xi_1^{\mathsf{T}}, \ldots, \xi_N^{\mathsf{T}}]^{\mathsf{T}}$, $\mathbf{X}_c = [{}^{W}\hat{\mathbf{X}}_1^{\mathsf{T}}, \ldots, {}^{W}\hat{\mathbf{X}}_n^{\mathsf{T}}]^{\mathsf{T}}$, and $\mathbf{X}_g = [{}^{W}\hat{\mathbf{B}}_1^{\mathsf{T}}, \ldots, {}^{W}\hat{\mathbf{B}}_n^{\mathsf{T}}]^{\mathsf{T}}$. We can solve the MLE of $\mathbf{p}$ by minimizing

$$\min_{\mathbf{p}} \omega(\mathbf{p})^{\mathsf{T}} \Sigma_{\omega}^{-1} \omega(\mathbf{p}) \tag{19}$$

where $\omega(\mathbf{p}) = [\omega_1^{\mathsf{T}}, \omega_2^{\mathsf{T}}, \omega_3^{\mathsf{T}}, \omega_4^{\mathsf{T}}]^{\mathsf{T}}$, $\Sigma_{\omega} = \mathrm{diag}(\Sigma_{\omega_1}, \Sigma_{\omega_2})$,

$\omega_1 = \begin{bmatrix} {}^{W}\hat{\mathbf{X}}_1 - {}^{W}\mathbf{X}_1 \\ \vdots \\ {}^{W}\hat{\mathbf{X}}_n - {}^{W}\mathbf{X}_n \end{bmatrix}, \psi_k = \begin{bmatrix} f({}^{W}\pi, \xi_k, {}^{W}\hat{\mathbf{X}}_1) - {}^{I_k}\bar{\mathbf{x}}_1 \\ \vdots \\ f({}^{W}\pi, \xi_k, {}^{W}\hat{\mathbf{X}}_n) - {}^{I_k}\bar{\mathbf{x}}_n \end{bmatrix}$,

$\omega_2 = [\cdots \psi_k^{\mathsf{T}} \cdots]^{\mathsf{T}}$, ψ_k is the reprojection error at stop k,

$\omega_3 = \begin{bmatrix} {}^{W}\hat{\mathbf{B}}_1 - {}^{W}\mathbf{B}_1 \\ \vdots \\ {}^{W}\hat{\mathbf{B}}_n - {}^{W}\mathbf{B}_n \end{bmatrix}, \rho_k = \begin{bmatrix} g(\eta, \xi_k, {}^{W}\hat{\mathbf{B}}_1) - {}^{G_k}\mathbf{B}_1 \\ \vdots \\ g(\eta, \xi_k, {}^{W}\hat{\mathbf{B}}_n) - {}^{G_k}\mathbf{B}_n \end{bmatrix}$,

$\omega_4 = [\cdots \rho_k^{\mathsf{T}} \cdots]^{\mathsf{T}}$, ρ_k is the metal ball center position estimation error at stop k, ${}^{I_k}\bar{\mathbf{x}}_j = \frac{\sum_{i=1}^{n} {}^{I_{k,i}}\mathbf{x}_j}{n}$, $\Sigma_{\omega_1} = \mathrm{diag}({}^{W}\Sigma_1 \ldots {}^{W}\Sigma_n, \Sigma_{\psi_1} \ldots \Sigma_{\psi_N})$, $\Sigma_{\psi_k} = \mathrm{diag}({}^{I_k}\Sigma_1 \ldots {}^{I_k}\Sigma_n)$, $\Sigma_{\omega_2} = \mathrm{diag}({}^{W}\Sigma_1 \ldots {}^{W}\Sigma_n, \Sigma_{\rho_1} \ldots \Sigma_{\rho_N})$, and $\Sigma_{\rho_k} = \mathrm{diag}({}^{G_k}\Sigma_1 \ldots {}^{G_k}\Sigma_n)$. ${}^{W}\Sigma_i$ and ${}^{I_k}\Sigma_i$ are obtained by direct measurement; ${}^{G_k}\Sigma_i$ is obtained from (17).

The problem can be solved by employing the LM algorithm. To show the covariance of η and ξ, we have Lemma 1 as follows and its proof in Appendix.

Lemma 1: Under the Gaussian noise assumption, the covariance matrix of η and ξ is given by

$$\Sigma_{\eta,\xi} = (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} \tag{20}$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are defined in (40).

D. Rigid Body Transformation Model Prediction Error

To verify the calibration results and the error propagation in Section V-C, we use the Euclidean distance between ${}^{G_k}\hat{\mathbf{B}}_i$ and ${}^{G_k}\mathbf{B}_i$ as the metric function to measure the model prediction error, where ${}^{G_k}\mathbf{B}_i$ is the measurement and ${}^{G_k}\hat{\mathbf{B}}_i$ is the estimator of ${}^{G_k}\mathbf{B}_i$. The metric function is defined by

$$\delta_{k,i} = h({}^{G_k}\hat{\mathbf{B}}_i, {}^{G_k}\mathbf{B}_i) = \|{}^{G_k}\hat{\mathbf{B}}_i - {}^{G_k}\mathbf{B}_i\|. \tag{21}$$

To compute (21), we estimate ${}^{G_k}\hat{\mathbf{B}}_i$ by calibrated parameters $\mathbf{q}_k^{\mathsf{T}} = \{\eta^{\mathsf{T}}, \xi_k^{\mathsf{T}}\}$ and measurements ${}^{W}\mathbf{B}_i$ according to (18). Lemma 2 shows the variance of $\delta_{k,i}$.

Lemma 2: Under the Gaussian noise assumption, the variance of $\delta_{k,i}$ is denoted by

$$\sigma^2_{\delta_{k,i}} = J_{h1}\,{}^{G_k}\hat{\Sigma}_i J_{h1}^{\mathsf{T}} + J_{h2}\,{}^{G_k}\Sigma_i J_{h2}^{\mathsf{T}} \tag{22}$$

where Jacobian matrices $J_{h1} = \frac{\partial h}{\partial {}^{G_k}\hat{\mathbf{B}}_i}$ and $J_{h2} = \frac{\partial h}{\partial {}^{G_k}\mathbf{B}_i}$.

Proof: Since $\delta_{k,i}$ is a function of ${}^{G_k}\hat{\mathbf{B}}_i$ and ${}^{G_k}\mathbf{B}_i$, the uncertainty of $\delta_{k,i}$ comes from their corresponding covariance matrices ${}^{G_k}\Sigma_i$ and ${}^{G_k}\hat{\Sigma}_i$. ${}^{G_k}\Sigma_i$ is given by (17); ${}^{G_k}\hat{\Sigma}_i$ is obtained by the forward propagation of error [45] under the first-order approximation

$${}^{G_k}\hat{\Sigma}_i = J_{g1}\Sigma_{\mathbf{q}_k} J_{g1}^{\mathsf{T}} + J_{g2}\,{}^{W}\Sigma_i J_{g2}^{\mathsf{T}} \tag{23}$$

where Jacobian matrices $J_{g1} = \frac{\partial g}{\partial \mathbf{q}_k}$, $J_{g2} = \frac{\partial g}{\partial {}^{W}\mathbf{B}_i}$, and g is the GPR parameterized function. ${}^{W}\Sigma_i$ is obtained from measurement and $\Sigma_{\mathbf{q}_k}$ is the covariance matrix of $\mathbf{q}_k$, which can be extracted by Lemma 1. Since ${}^{G_k}\hat{\mathbf{B}}_i$ and ${}^{G_k}\mathbf{B}_i$ have no correlation, the overall variance of $\delta_{k,i}$ in (22) can be obtained by addition of their uncertainties. ■

To compute the average error for each metal ball center position, we first estimate $\{(\delta_{k,i}, \sigma^2_{\delta_{k,i}})|k=1,\ldots,N\}$, and then take the average of them by $\delta_i = \sum_{k=1}^{N} \delta_{k,i}/N$. Finally, the expected value δ_i and its variance $\sigma^2_{\delta_i}$ can be estimated by using sample mean and sample variance.

VI. Biomodal Surface and Subsurface Mapping

After the calibration, we obtain the frame mapping relationship between the camera and the GPR. This spatial relationship alone cannot help us to fuse data from both modalities because we need to address the issue of synchronization.

Recall that camera images are triggered by time and GPR scans are indexed by the wheel encoder pluses based on a preset fixed distance traveled. To facilitate the synchronization of the two data streams, we design ALs [see Fig. 6(b)] that are made of colored patterned metal plates and clearly visible to both sensors, which enable us to align surface and subsurface structure from two different sensor modalities.

We also design a data collection procedure [see Fig. 6(a)]. We first place ALs on the survey area and make sure that those ALs are evenly spaced out and remain fixed on the ground. The number of ALs is adjustable in our setting, but we suggest to use at least two: one at the starting point and the other at the ending point. Next, we push the sensing suite to traverse ALs by following the preplanned survey trajectory. During the

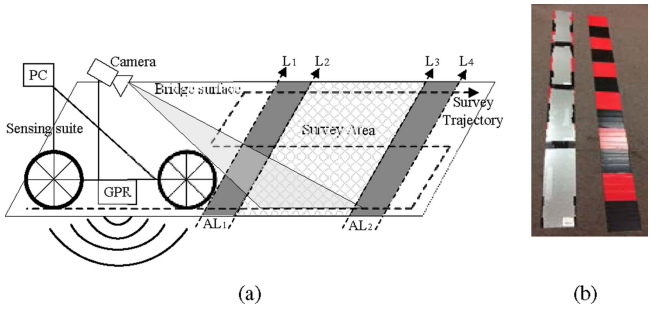


Fig. 6. (a) Illustration of the data collection procedure. (b) Artificial landmarks. Black and red colored side is the upper side, whereas the metal side is the downside (best viewed in color).

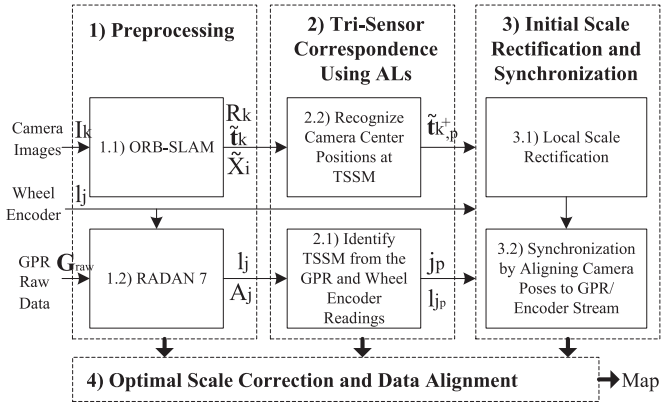


Fig. 7. Software diagram.

procedure, we ensure that the sensing suite traverses the edges of ALs on both sides along the trajectory. Fig. 6(a) shows an example of the setup: four lines $\{\mathbf{L}_1, \mathbf{L}_2, \mathbf{L}_3, \mathbf{L}_4\}$ indicating the four edges generated by both sides of the two ALs.

Employing ALs makes data collection procedure more efficient and flexible if comparing to a conventional approach, which relies on manually painted lines or grids on the survey area. The conventional approach also requires the GPR to painstakingly move along a preset linear trajectory, which is time consuming and labor intensive. In our design, we only need to place ALs on the ground without requiring precise positioning or strict linear scan motion of the sensing suite.

To formulate the synchronization-challenged sensor fusion problem, we follow the assumptions and notations defined in Section V, except for the modifications listed in the following.

1) $\{C_k\}$ denotes the camera coordinate system at time k, and I_k denotes the camera image at time k, where $k = 1, 2, \ldots, N_k$. Here, we change index variable k from the previous mechanic stop index to time index.
2) $\{W\}$ is coinciding with $\{C_1\}$. To simplify the notations, we will ignore the reference frame (left superscript) if describing a point w.r.t. $\{W\}$. For example, $\mathbf{X} \in \mathbb{R}^3$ represents a certain point w.r.t. $\{W\}$.
3) $\mathbf{G}_{\text{raw}}$ denotes all the GPR raw data.

Now, we define the synchronization problem as follows.

Definition 2: Given camera image sequences I_k and GPR raw data $\mathbf{G}_{\text{raw}}$ with wheel encoder data l_j, where $k = 1, 2, \ldots, N_k$ and $j = 1, 2, \ldots, m$, determine the 3-D metric reconstruction map including camera and GPR poses, 3-D landmarks, and GPR data.

Fig. 7 summarizes our software diagram in four steps. We will detail them in the following sections.

A. Preprocessing

Camera images and GPR scans are preprocessed to obtain initial reconstruction information in its individual modality. For camera images, we apply the popular ORB-SLAM2 [25] to estimate camera poses and 3-D landmarks at each frame. Given sequential camera images I_k, we obtain camera poses $\{\mathbf{R}_k, \tilde{\mathbf{t}}_k\}$ w.r.t. $\{W\}$ at time k, where $\mathbf{R}_k$ is the camera rotation matrix, $\tilde{\mathbf{t}}_k \in \mathbb{R}^3$ is the translation vector, and $k = 1, 2, \ldots, N_k$. The ORB-SLAM2 algorithm also provides 3-D landmark positions $\tilde{\mathbf{X}}_i \in \mathbb{R}^3$, where index i denotes the ith 3-D landmark, and $i = 1, 2, \ldots, N_i$. Note that this is the outcome of the monocular vSLAM; all 3-D information is up to scale. As a convention in this section, we use the tilde $\tilde{}$ to indicate variables in 3-D space that are up-to-scale. Therefore, $\tilde{\mathbf{t}}_k$ and $\tilde{\mathbf{X}}_i$ are up to scale, whereas $\mathbf{t}_k$ and $\mathbf{X}_i$ are defined in the metric scale. For GPR raw data $\mathbf{G}_{\text{raw}}$ and wheel encoder reading l_j, we use GSSI software RADAN 7 to export the GPR A-scans with accumulated travel length $\{A_j, l_j\}$, where $j = 1, 2, \ldots, m$.

As a result, we obtain camera readings $\{\mathbf{R}_k, \tilde{\mathbf{t}}_k, \tilde{\mathbf{X}}_i\}$ and GPR/encoder readings $\{A_j, l_j\}$ after the preprocessing step. For now, we do not know wheel encoder readings corresponding to each camera frame yet, which is the focus of the next step.

B. Signal Correspondence Using ALs

From Section V, we know the relative pose between the camera and the GPR. Let us define AL anchor points to be the points on the AL edges created by the intersection between the projected camera/GPR ground trajectory and AL edges. The three sensor synchronization moment (TSSM) is the moment that the camera center, GPR coordinate system origin, and AL anchor point are collinear as a line perpendicular to the ground (see Fig. 8). If we recognize this moment across all sensor modalities, the synchronization problem can be solved.

1) Identify TSSM From the GPR and Wheel Encoder Readings: When the GPR origin is traversing an AL, the metal plate generates a strong response as described in Section IV-E as shown in Fig. 8. The TSSMs occur on the AL edges. Therefore, we utilizes the GPR response to the ALs' edges at TSSMs to determine the correct scan indexes in order to perform sensor fusion. Let the number of detected edges be M for all ALs in the setting. After searching the edge scan indexes from the GPR image, we apply (7) to obtain corrected wheel encoder reading l_{j_p} and the corrected GPR scan indexes j_p, where $p = 1, 2, \ldots, M$.

2) Recognize Camera Center Positions at TSSM: For now, we only have an up-to-scale 3-D reconstruction from the ORB-SLAM2 results. We need to obtain TSSM camera center positions in this coordinate system. Note this camera pose is a virtual pose, which means it does not correspond to an actual image due

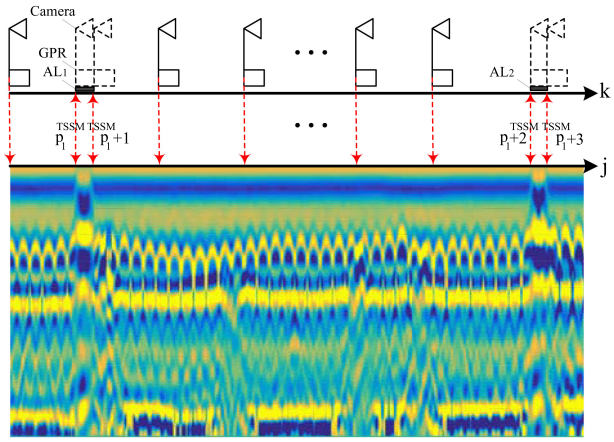


Fig. 8. ALs help us synchronize camera image stream and GPR/encoder data streams. Four TSSMs are shown here. Camera poses represented by small triangles and GPR poses represented by small rectangles are displayed on the top of the radargram. The poses drawn in dashed lines are virtual poses corresponding to TSSMs (best viewed in color).

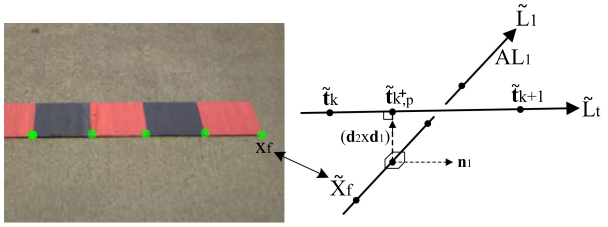


Fig. 9. Left: AL edge points in a 2-D camera image. Right: two skew lines $\tilde{\mathbf{L}}_1$ and $\tilde{\mathbf{L}}_t$ are used to compute camera center position $\tilde{\mathbf{t}}_{k^+,p}$ (best viewed in color).

to the discretized time across frames. However, this pose can help us synchronize all sensors later. The visual poses drawn in dashed line in Fig. 8 give an example about the geometry relationship between sensors at TSSM.

Fig. 9 illustrates the process of identifying the camera center position for the virtual pose. The high contrast pattern in AL allows us to recognize points on the edge (green points in the figure) across multiple frames. Based on the known camera poses, we obtain their 3-D points. Let $\{\mathbf{x}_f \leftrightarrow \tilde{\mathbf{X}}_f | f \in \mathscr{L}_1\}$ be the corresponding 2-D and 3-D points where $\mathscr{L}_1$ is the index set containing all edge points. These 3-D points $\tilde{\mathbf{X}}_f$ allows us to establish AL edge line $\tilde{\mathbf{L}}_1 : \tilde{\mathbf{X}} = \mathbf{p}_1 + t_1\mathbf{d}_1$, where t_1 is a parameter, $\mathbf{p}_1 = \text{mean}(\tilde{\mathbf{X}}_f)$, and $\mathbf{d}_1$ is the singular vector corresponding to the largest singular value after performing singular value decomposition (SVD) on normalized points $[\tilde{\mathbf{X}}_1 - \mathbf{p}_1, \ldots, \tilde{\mathbf{X}}_f - \mathbf{p}_1, \ldots]^\mathsf{T}$.

From images taken, we identify two image indexes k and $k+1$ representing two immediate camera poses before and after the crossing between projection of $\tilde{\mathbf{L}}_\mathbf{t}$ on the ground and $\tilde{\mathbf{L}}_\mathbf{1}$. The camera center positions at k and $k+1$ are $\tilde{\mathbf{t}}_k$ and $\tilde{\mathbf{t}}_{k+1}$, respectively. The camera trajectory between k and $k+1$ can be approximated by a line $\tilde{\mathbf{L}}_t : \tilde{\mathbf{X}} = \mathbf{p}_2 + t_2\mathbf{d}_2$, where t_2 is a parameter, $\mathbf{p}_2 = \tilde{\mathbf{t}}_k$, and $\mathbf{d}_2 = \tilde{\mathbf{t}}_k - \tilde{\mathbf{t}}_{k+1}$.

Finding the skew line between $\tilde{\mathbf{L}}_1$ and $\tilde{\mathbf{L}}_t$ allows us to obtain the camera center positions $\tilde{\mathbf{t}}_{k^+,p}$ at TSSM, where the combo subscripts k^+, p indicate this is slightly after time k and it is corresponding to the pth AL edge. We have

$$\tilde{\mathbf{t}}_{k^+,p} = \mathbf{p}_2 + \frac{(\mathbf{p}_1 - \mathbf{p}_2)\cdot \mathbf{n}_1}{\mathbf{d}_2 \cdot \mathbf{n}_1} \cdot \mathbf{d}_2 \tag{24}$$

where $\mathbf{n}_1 = \mathbf{d}_1 \times (\mathbf{d}_2 \times \mathbf{d}_1)$.

C. Initial Scale Rectification and Synchronization

Now, we can recover the true scale by using TSSM correspondence.

1) Local Scale Rectification: When the sensing suite traverses two adjacent ALs, we identify two virtual camera poses corresponding to the leading edge of each AL. If the first edge index is p_1, then the second edge index is $p_1 + 2$ due to the adjacency (see Fig. 8). Say that the corresponding camera frame indexes are k_1 and k_2. Then, the camera center positions of the two virtual poses are $\tilde{\mathbf{t}}_{k_1^+,p_1}$ and $\tilde{\mathbf{t}}_{k_2^+,p_1+2}$. The corresponding distances extracted from encoder readings are $l_{j_{p_1}}$ and $l_{j_{p_1+2}}$, respectively. Then, a local scale ratio can be obtained as the ratio between the distance from the encoder and the distance from camera poses as (25), shown at the bottom of this page, where $||\cdot||$ is l_2 norm. Then, the scale of all $\tilde{\mathbf{t}}_k$ and $\tilde{\mathbf{X}}_i$ at frames k between k_1+1 and k_2 can be recovered by

$$\mathbf{t}_k = s\tilde{\mathbf{t}}_k, \quad \mathbf{X}_i = s\tilde{\mathbf{X}}_i. \tag{26}$$

For $\tilde{\mathbf{t}}_k$ and $\tilde{\mathbf{X}}_i$ before the first virtual camera pose or after the last virtual pose, we can also use the closest local scale ratio to rectify them. Note that we use a local scale ratio instead of a global scale ratio established by the first and the last ALs because monocular vSLAM may not have a constant scale due to scale drift. Using a local scale to correct the affected poses and 3-D landmarks can alleviate the issue. Also, this is not a precise scale recovery because the encoder readings may be inaccurate due to skids and the scale drift variation may be big. We will handle it in Section VI-D.

2) Synchronizing Camera Poses to GPR/Encoder Data: With the scale rectified, we align camera poses with the GPR/encoder data streams through distance matching. Let $\mathbf{t}'_k$ be GPR frame origin for the corresponding camera center position $\mathbf{t}_k$. The fixed frame mapping relationship is

$$\mathbf{t}'_k = \mathbf{R}_{\text{ex}}\mathbf{t}_k + \mathbf{t}_{\text{ex}} \tag{27}$$

where extrinsic parameters $\{\mathbf{R}_{\text{ex}}, \mathbf{t}_{\text{ex}}\}$ are obtained in Section V. Define the distance traveled from the first edge of the first AL to the current pose k_c to be $d_0(k_c)$. Say that k_0 is the index of the camera/GPR frame right before the first edge, and $\mathbf{t}_{k_0^+,1}$ is the

$$s = \frac{l_{j_{p_1+2}} - l_{j_{p_1}}}{||\tilde{\mathbf{t}}_{k_1+1} - \tilde{\mathbf{t}}_{k_1^+,p_1}|| + \sum_{k=k_1+1}^{k_2-1} ||\tilde{\mathbf{t}}_{k+1} - \tilde{\mathbf{t}}_k|| + ||\tilde{\mathbf{t}}_{k_2^+,p_1+2} - \tilde{\mathbf{t}}_{k_2}||} \tag{25}$$

camera center of the first virtual pose right above the first edge and used as the starting point of the inspection. Then, we have

$$d_0(k_c) = ||\mathbf{t}'_{k_0+1} - \mathbf{t}'_{k_0^+,1}|| + \sum_{k=k_0+1}^{k_c-1} ||\mathbf{t}'_{k+1} - \mathbf{t}'_k||. \quad (28)$$

Note that we do not have ˜ over variables because they are in metric space.

Define $d_l(k_c)$ as the corresponding cumulative distance traveled from the first edge of the first AL to the current GPR scan index j_{k_c}, and we have

$$d_l(k_c) = l_{j_{k_c}} - l_0 \quad (29)$$

where $l_{j_{k_c}}$ is the encoder reading at the current GPR scan index, and l_0 is the encoder reading corresponding to the first virtual pose $\mathbf{t}'_{k_0^+,1}$.

Based on (28) and (29), we are able to compute $d_0(k)$ for each camera/GPR pose k, and its corresponding $d_l(k)$ from wheel encoder reading.

D. Optimal Scale Correction and Data Alignment

Now, we further synchronize sensor readings and perform metric reconstruction using an optimization framework. We formulate a constrained optimization problem as follows. Let the estimated parameters be 3-D landmarks $\mathscr{X} = \{\mathbf{X}_i | i = 1, \ldots, N_i\}$, camera orientations $\mathscr{R} = \{\mathbf{R}_k | k = 1, \ldots, N_k\}$, and camera center positions $\mathscr{T} = \{\mathbf{t}_k | k = 1, \ldots, N_k\}$. The cost function for camera reprojection errors is given by

$$C(\mathscr{X}, \mathscr{R}, \mathscr{T}) = \sum_{k=1}^{N_k} \sum_{i \in \mathscr{S}_k} \|\mathbf{x}_{i,k} - \hat{\mathbf{x}}_{i,k}\|^2_{\Sigma_{i,k}} \quad (30)$$

where $\mathscr{S}_k$ denotes the set containing all indexes of points visible by camera at time k, $\mathbf{x}_{i,k}$ is the image observation of $\mathbf{X}_i$ at time k, $\hat{\mathbf{x}}_{i,k} = f(\mathbf{R}_k, \mathbf{t}_k, \mathbf{X}_i)$ is the camera projection function, $\Sigma_{i,k}$ is the covariance of $\mathbf{x}_{i,k}$, and $\| \cdot \|_\Sigma$ denotes Mahalanobis distance. It is worth noting that points $\mathbf{x}_{i,k}$ are the surviving inline set from ORB-SLAM2 pose graph.

At each camera/GPR frame, we incorporate encoder readings to capture the traveled distance. Note that encoder error accumulates linear to the distance traveled. We verify distances traveled between adjacent camera/GPR poses and formulate the following objective function by considering relative error:

$$F(\mathscr{X}, \mathscr{T}) = \sum_{k=2}^{N_k} \left\| \frac{[d_0(k) - d_0(k-1)] - [d_l(k) - d_l(k-1)]}{d_l(k) - d_l(k-1)} \right\|. \quad (31)$$

In addition, virtual poses at AL edges provide more constraints to this problem that can be used as penalty functions in the objective function. For each rectified camera center $\mathbf{t}_{k^+,p}$ for the virtual pose, we find its corresponding virtual GPR pose at $\mathbf{t}'_{k^+,p}$ using (27). Similar to (28) and (29), we define travel distance function $d_v(p)$ as

$$d_v(p) = ||\mathbf{t}'_{k_0+1} - \mathbf{t}'_{k_0^+,1}|| + \sum_{k=k_0+1}^{k_p-1} ||\mathbf{t}'_{k+1} - \mathbf{t}'_k|| + ||\mathbf{t}'_{k_p^+,p} - \mathbf{t}'_{k_p}|| \quad (32)$$

where k_p is the index of the camera/GPR frame right before the pth AL and $d_v(1) = 0$. Then, we have

$$G(\mathscr{X}, \mathscr{T}) = \sum_{p=1}^{M/2} \left\| \frac{[d_v(2p) - d_v(2p-1)] - [l_{j_{2p}} - l_{j_{2p-1}}]}{l_{j_{2p}} - l_{j_{2p-1}}} \right\| \quad (33)$$

where $l_{j_{2p}}$ and $l_{j_{2p-1}}$ are the corrected encoder readings for each sequential virtual poses at the beginning edge and the ending edge of the AL. Finally we formulate the optimization problem as

$$\min_{\mathscr{X}, \mathscr{R}, \mathscr{T}} C(\mathscr{X}, \mathscr{R}, \mathscr{T}) + \alpha \cdot F(\mathscr{X}, \mathscr{T}) + \beta \cdot G(\mathscr{X}, \mathscr{T}) \quad (34)$$

where α and β are nonnegative weighting scalars. In (34), $F(\mathscr{X}, \mathscr{T})$ is a soft constraint due to potential synchronization errors, and $G(\mathscr{X}, \mathscr{T})$ is a hard constraint because the AL correspondence at TSSM must be strictly preserved. Therefore, β has a higher value than α. We have to adjust α and β to solve (34). The optimization problem can be solved by first using a small positive weight for α and β, and then applying any nonlinear optimization solver, e.g., LM. Then, we gradually increase α and β and use the previous solution as the initial solution to solve (34) iteratively. Finally, we obtain the solution as β is sufficiently large and the residual is converged. After estimating $\mathscr{X}, \mathscr{R}$, and $\mathscr{T}$, we repeat the synchronization procedure in Section VI-C2 to further remove errors caused by scale drift, and reoptimize to improve accuracy.

VII. Experiments

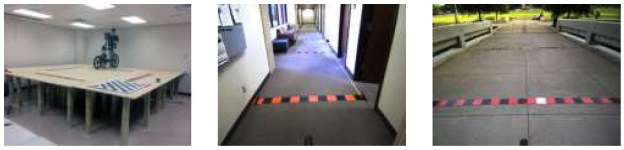
The proposed system has been validated in physical experiments. We first show the results of calibration and then verify our sensor fusion algorithm.

For our GPR, the horizontal sample rate for the wheel encoder is 390 pulses per meter, the two-way travel time of the radar signal is 8 ns, the sample rate for the GPR control unit is 1024 samples/scan, and the dielectric constant in air is 1. The width of each AL is $w = 0.152$ m as measured by a ruler. The offset for each AL is $\delta_{\text{tran}} = 0.031$ m using the method in Section IV-E.

A. Calibration Experiment

1) Calibration Experiment Setup and Results: Fig. 2(b) shows the sensing suite and the calibration rig design. The intrinsic camera parameters are calibrated using camera calibration toolbox for MATLAB [47]. The radius of the metal ball is 19.0 mm and vertical height h_i is 419.0 mm. The side length of each square in the mirror checkerboard is 128.0 mm, and the side length of each square in ball checkerboard is 56.0 mm.

The calibration data collection follows the procedure mentioned at the beginning of Section V with $n = 80$ and $N = 20$ in settings. We collect a calibration set and a test set for cross validation. The resulting relative pose between the camera and the GPR from the calibration set is ${}^G_C\mathbf{t} = [0, -60, 1180]^\mathsf{T}$mm



(a) (b) (c)

Fig. 10. (a) Physical experiment in the artificial platform. (b) Field experiment in the corridor. (c) Field experiment in the bridge.

and $[\theta_x, \theta_y, \theta_z]^\mathsf{T} = [-1.9909, 0.0270, 0.0180]^\mathsf{T}$ rad, where $(\theta_x, \theta_y, \theta_z)$ is the Euler angle representation of ${}^{G}_{C}\mathbf{R}$.

To validate our calibration results, we utilize the Euclidean distance between the estimated metal ball center positions and the measured metal ball center positions ${}^{W}\mathbf{B}_i$ as the error metric

$$e_i = \| {}^{W}_{C_k}\mathbf{T} {}^{C_k}_{G_k}\mathbf{T} {}^{G_k}\mathbf{B}_i - {}^{W}\mathbf{B}_i \| \tag{35}$$

where ground truth ${}^{W}\mathbf{B}_i$ is from the test set and directly measured using the ruler, ${}^{W}_{C_k}\mathbf{T}$ and ${}^{C_k}_{G_k}\mathbf{T}$ are estimated from the calibration set, and ${}^{G_k}\mathbf{B}_i$ is computed from B_i and h_i in the test set. The results show that the mean error and standard deviation (SD) are 8.0 mm and 5.0 mm, respectively. Considering the fact that the GPR signal wave length is 18.75 cm, the results are satisfying.

2) Model Prediction Error and Uncertainty Analysis: Now, we evaluate if our uncertainty analysis in Section V-D can capture the prediction error of the calibrated model. We illustrate the model prediction error δ_i and the predicted SD σ_{δ_i} for the 80 testing samples. More specifically, the measurement errors for metal ball center position measurements have a variance of 8.0 mm^2 in each dimension. This is usually caused by radar accuracy and structural deformation underweight. Besides, the variance for points on GPR and camera images are 1 pixel^2. The results agree with our analysis as 68.75% errors fall in the $1-\sigma$ range of the calibrated model prediction.

B. Bimodal Surface and Subsurface Mapping

1) Experiment Setup and Results: To validate our result, we conduct two kinds of physical experiments: 1) controlled indoor experiments in an artificial platform to verify the accuracy; and 2) field experiments to test the performance.

For 1), we build the artificial platform by using wooden piles and boards as shown in Fig. 10(a). The platform is a 5.5×5.5 m^2 planar surface with a height of 0.9 m above the ground. This artificial platform enables us to collect accurate ground truth for further validation (see Sections VII-B2 and VII-B3).

During data collection, we push the sensing suite to traverse ALs by following the preplanned survey trajectory and obtain both camera and GPR data. This setup provides highly accurate ground truth by directly measuring the distance between the ALs. We have ten preplanned 2.830-m trajectories. We repeat 10 independent trials on each trajectory with a total of 100 trials.

For 2), we perform both indoor and outdoor environments at Texas A&M University [see Fig. 10(b) and (c)]. The indoor experiment covers a corridor on the third floor of the HRBB building and the outdoor experiment covers a bridge deck at the Ernest Langford architecture center. The indoor tests are repeated for 4 independent trials and each trial trajectory length is 3.048 m; the outdoor tests are repeated for 14 independent trials and each trial trajectory length is 4.084 m.

Fig. 1 illustrates the 3-D reconstruction result for the experiments in the bridge (other results can be found in the attached video file). The green points indicate visual 3-D landmarks, and magenta lines indicate ALs. The camera poses are shown at the top of the 3-D map, whereas the synchronized GPR poses are aligned at the bottom. Besides, GPR scans are attached to GPR positions following the GPR trajectory. In all experiments, our algorithms are able to synchronize these data streams to create successful 3-D metric reconstruction.

2) Accuracy Test for the Travel Distance Error: To validate if our algorithm can improve the accuracy for 3-D reconstruction, we adopt the end-to-end GPR travel distance error between the first and the last ALs as the metric since it is a good approach to evaluate the accuracy of the overall geometry of the mapped environment by using an accurate external measurement for a known relative displacements of faraway poses [48], [49]. For the field experiment, we manually measure the distance d_{GT} that GPR traveled from the first AL's edge ($p = 1$) to the last AL's edge ($p = M$) when collecting data, and treat it as the ground truth. For the artificial platform, we obtain the ground truth (d_{GT}) with high accuracy in advance as mentioned in the setup.

We compare the following three approaches:

1) encoder-camera-GPR sensor fusion;
2) encoder measurement only;
3) nonsynchronized encoder-camera sensor fusion.

First, we define the error metric for our sensor fusion algorithm as

$$e_{\text{ECG}} = \left| \|\mathbf{t}'_{k^+,M} - \mathbf{t}'_{k^+,1}\| - d_{\text{GT}} \right| \tag{36}$$

where $\mathbf{t}'_{k^+,1}$ denotes the first GPR virtual pose and $\mathbf{t}'_{k^+,M}$ denotes the last GPR virtual pose on the trajectory. Second, we define the corresponding error metric for the encoder measurement only as

$$e_E = \left| [l_{j_M} - l_{j_1}] - d_{\text{GT}} \right| \tag{37}$$

where l_{j_1} denotes the first AL's edge and l_{j_M} denotes the last AL's edge. Third, we define the corresponding error metric for the nonsynchronized encoder-camera sensor fusion as

$$e_{\text{EC}} = \left| \|\mathbf{t}^*_{k^+,M} - \mathbf{t}^*_{k^+,1}\| - d_{\text{GT}} \right| \tag{38}$$

where $\mathbf{t}^*_{k^+,1}$ denotes the first GPR virtual pose, and $\mathbf{t}^*_{k^+,M}$ denotes the last GPR virtual pose. Here, we use the nonsynchronized encoder-camera fusion as the comparison to show the case without the help from the GPR for synchronization. The nonsynchronized encoder-camera fusion is conducted by using overall distance from encoder readings to rectify scale for camera trajectory as a whole.

Table I shows the error metrics for three approaches in all experiments. The No. indicates the number index of each trial, so the corridor and bridge results are single trials; the platform results are average of ten trials. Detailed surface feature and subsurface maps are included in the attached video file.The results show that our algorithm improves the accuracy for 3-D reconstruction because its error is consistently less than that of

TABLE I
ERROR METRICS FOR END-TO-END TRAVEL DISTANCES IN THREE SCENARIOS: THE PLATFORM, THE CORRIDOR, AND THE BRIDGE

Artificial Platform

No.	e_{ECG} (m)	e_E(m)	e_{EC}(m)
1-10	**0.020**	0.026	0.103
11-20	**0.009**	0.025	0.074
21-30	**0.010**	0.023	0.065
31-40	**0.013**	0.027	0.110
41-50	**0.017**	0.033	0.090
51-60	**0.006**	0.022	0.076
61-70	**0.009**	0.029	0.100
71-80	**0.009**	0.025	0.078
81-90	**0.011**	0.023	0.069
91-100	**0.012**	0.028	0.104
Mean	**0.012**	0.026	0.087
Std.	0.004	**0.003**	0.016

Indoor Corridor

No.	e_{ECG}(m)	e_E(m)	e_{EC}(m)
1	**0.010**	0.037	0.066
2	**0.008**	0.034	0.047
3	**0.010**	0.032	0.066
4	**0.012**	0.026	0.043
Mean	**0.010**	0.032	0.056
Std.	**0.002**	0.005	0.012

Outdoor Bridge

No.	e_{ECG} (m)	e_E(m)	e_{EC}(m)
1	**0.009**	0.077	0.011
2	0.009	0.066	**0.007**
3	0.008	0.074	**0.006**
4	**0.008**	0.074	0.022
5	**0.004**	0.069	0.032
6	**0.006**	0.082	0.043
7	**0.003**	0.067	0.053
8	**0.009**	0.053	0.040
9	**0.006**	0.048	0.250
10	**0.007**	0.069	0.235
11	**0.005**	0.074	0.007
12	**0.010**	0.059	0.023
13	**0.001**	0.025	0.148
14	**0.011**	0.053	0.160
Mean	**0.007**	0.064	0.074
Std.	**0.003**	0.015	0.086

Bold entities represent the best results in comparison.

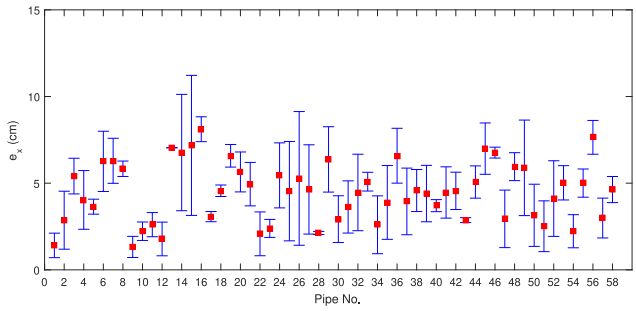


Fig. 11. Center position error e_x for each pipe [9]. Each red marker position is the mean, and each blue vertical segment is its corresponding $1-\sigma$ range. The overall average center position error is 4.47 cm.

As shown in Fig. 11, we number each pipe by its index in the *X*-axis, and illustrate the statistical results of e_x for each pipe in the *Y*-axis. Each red marker position is the mean, and each blue vertical segment is its corresponding $1-\sigma$ range. The overall average center position error is 4.47 cm. This is sufficiently accurate considering the wavelength of the GPR, and hence the overall design is successful.

the counterparts in the experiments. It is worth mentioning that the results of nonsynchronized encoder-camera sensor fusion is even worse than that of the encoder measurement because the encoder and the camera are not synchronized (without the help from the GPR).

3) Accuracy Test for the Underground Pipe Mapping Error: Besides validating the end-to-end GPR travel distance error in Section VII-B2, we further verify our bimodal surface and subsurface mapping results by mapping subsurface cylindrical pipes.

In the experiment setup, the checkerboard attached to the surface [see Fig. 10(a)] is used to build the world coordinate for us to validate our experimental results. We drill holes on the corners on the checkerboard so that we are able to align the 3-D space under surface to the world coordinate; hence, we can further utilize that checkerboard to build the ground truth of all the pipelines under the surface.

To build ground truth, we first place pipes under the artificial platform surface and manually measure the center positions of as the ground truth using a tape measure with 1.59-mm accuracy. Then, we estimate the center positions of each cylinder landmark by using our 3-D reconstruction results and compare with the ground truth to validate our system. Since the detailed pipe recognition and reconstruction is presented in our previous papers [8], [9] and is not the focus of this article, we omit the detail here.

Let e_x denote the Euclidean distance of the center position between the estimation and the ground truth. We directly apply e_x for each pipe as the mapping accuracy metric. Fig. 11 shows the result. We totally test 58 pipes in our 30 experimental trials.

VIII. CONCLUSION

In this article, we proposed a system for both surface and subsurface infrastructure inspection using a multimodal sensing suite. First, we developed a calibration rig to estimate the relative pose of a GPR and a camera. We modeled the camera projection with mirror reflection transformation and the GPR imaging process. We estimated camera and mirror poses from camera images and extracted hyperbolas in the GPR image to recover metal ball coordinates. The MLE was employed to estimate the rigid body transformation between the two sensors. We provided the closed-form error analysis for our calibration models. Second, we designed a data collection scheme by using ALs to assist the synchronization between camera images (temporally evenly spaced) and GPR/encoder data (spatially evenly spaced). We identified synchronization events created by ALs and use them as inputs to synchronize sensory inputs. We used the GPR/wheel encoder readings to rectify the scale for monocular vSLAM and then employed pose graph optimization by considering synchronization constraints. We implemented our algorithm and tested it in physical experiments. Our system and algorithm have successfully achieved sensor calibration, data synchronizations, and 3-D metric reconstruction.

In the future, we will develop cross-modality scene reconstruction and recognition algorithms for surface and subsurface inspection tasks. We will also consider incorporating other sensors, such as LIDARs and inertial sensors. We will allow GPR to move long arbitrary trajectory when developing new reconstruction algorithms. More field tests and experiments will be reported as they emerge.

$$J_\omega^\mathsf{T}\Sigma_\omega^{-1}J_\omega = \begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix} = \begin{bmatrix} J_\eta^\mathsf{T}{}^\rho\Sigma J_\eta & J_\eta^\mathsf{T}{}^\rho\Sigma J_{\xi 2} & \mathbf{0} & \mathbf{0} & J_\eta^\mathsf{T}{}^\rho\Sigma J_{x2} \\ J_{\xi 2}^\mathsf{T}{}^\rho\Sigma J_\eta & J_{\xi 1}^\mathsf{T}{}^\psi\Sigma J_{\xi 1} + J_{\xi 2}^\mathsf{T}{}^\rho\Sigma J_{\xi 2} & J_{\xi 1}^\mathsf{T}{}^\psi\Sigma J_\pi & J_{\xi 1}^\mathsf{T}{}^\psi\Sigma J_{x1} & J_{\xi 2}^\mathsf{T}{}^\rho\Sigma J_{x2} \\ \mathbf{0} & J_\pi^\mathsf{T}{}^\psi\Sigma J_{\xi 1} & J_\pi^\mathsf{T}{}^\psi\Sigma J_\pi & J_\pi^\mathsf{T}{}^\psi\Sigma J_{x1} & \mathbf{0} \\ \mathbf{0} & J_{x1}^\mathsf{T}{}^\psi\Sigma J_{\xi 1} & J_{x1}^\mathsf{T}{}^\psi\Sigma J_\pi & {}^C\Sigma + J_{x1}^\mathsf{T}{}^\psi\Sigma J_{x1} & \mathbf{0} \\ J_{x2}^\mathsf{T}{}^\rho\Sigma J_\eta & J_{x2}^\mathsf{T}{}^\rho\Sigma J_{\xi 2} & \mathbf{0} & \mathbf{0} & {}^G\Sigma + J_{x2}^\mathsf{T}{}^\rho\Sigma J_{x2} \end{bmatrix} \tag{40}$$

APPENDIX
PROOF OF LEMMA 1

Proof: From the first-order approximation of error backward propagation [45], we obtain the covariance matrix of $\mathbf{p}$ by

$$\Sigma_\mathbf{p} = \left(J_\omega^\mathsf{T}\Sigma_\omega^{-1}J_\omega\right)^{-1} \tag{39}$$

where $\Sigma_\omega^{-1} = \left[\begin{smallmatrix}\Sigma_{\omega_1}^{-1} & \mathbf{0}\\ \mathbf{0} & \Sigma_{\omega_2}^{-1}\end{smallmatrix}\right]$ and Jacobian matrices

$$J_\omega = \frac{\partial\omega(\mathbf{p})}{\partial\mathbf{p}} = \begin{bmatrix}\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0}\\ \mathbf{0} & J_{\xi 1} & J_\pi & J_{x1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}\\ J_\eta & J_{\xi 2} & \mathbf{0} & \mathbf{0} & J_{x2}\end{bmatrix}$$

$$J_\eta = \left[\begin{smallmatrix}\frac{\partial\rho_1^\mathsf{T}}{\partial\eta}, & \cdots, & \frac{\partial\rho_N^\mathsf{T}}{\partial\eta}\end{smallmatrix}\right]^\mathsf{T}, J_\pi = \left[\begin{smallmatrix}\frac{\partial\psi_1^\mathsf{T}}{\partial{}^W\pi}, & \cdots, & \frac{\partial\psi_N^\mathsf{T}}{\partial{}^W\pi}\end{smallmatrix}\right]^\mathsf{T}$$

$$J_{\xi 1} = \operatorname{diag}\left(\frac{\partial\psi_1}{\partial\xi_1}, \ldots, \frac{\partial\psi_N}{\partial\xi_N}\right), J_{\xi 2} = \operatorname{diag}\left(\frac{\partial\rho_1}{\partial\xi_1}, \ldots, \frac{\partial\rho_N}{\partial\xi_N}\right)$$

$$J_{x1} = \left[\begin{smallmatrix}\frac{\partial\psi_1^\mathsf{T}}{\partial\mathbf{X}_c}, & \cdots, & \frac{\partial\psi_N^\mathsf{T}}{\partial\mathbf{X}_c}\end{smallmatrix}\right]^\mathsf{T}, J_{x2} = \left[\begin{smallmatrix}\frac{\partial\rho_1^\mathsf{T}}{\partial\mathbf{X}_g}, & \cdots, & \frac{\partial\rho_N^\mathsf{T}}{\partial\mathbf{X}_g}\end{smallmatrix}\right]^\mathsf{T}.$$

To simplify the notations, we denote $\Sigma_{\omega_1}^{-1} = \left[\begin{smallmatrix}{}^C\Sigma & \mathbf{0}\\ \mathbf{0} & {}^\psi\Sigma\end{smallmatrix}\right]$, and $\Sigma_{\omega_2}^{-1} = \left[\begin{smallmatrix}{}^G\Sigma & \mathbf{0}\\ \mathbf{0} & {}^\rho\Sigma\end{smallmatrix}\right]$, and derive (40) at the top of next page. Finally, we solve $\Sigma_{\eta,\xi}$ through $\Sigma_\mathbf{p}$ by applying the blockwise matrix inversion to (40), shown at the top of this page. ■

ACKNOWLEDGMENT

The authors would like to thank H. Cheng, B. Li, S. Yeh, A. Angert, A. Kingery, D. Wang, and Y. Ou for their inputs and contributions to the NetBot Lab at Texas A&M University.

REFERENCES

[1] A. Elatta, L. P. Gen, F. L. Zhi, Y. Daoyuan, and L. Fei, "An overview of robot calibration," *Inf. Technol. J.*, vol. 3, no. 1, pp. 74–78, Jan. 2004.

[2] S. Ross, *Introduction to Probability Models*, 9th ed. New York, NY, USA: Academic, 2007.

[3] B. Siciliano and O. Khatib, *Springer Handbook of Robotics*. New York, NY, USA: Springer, 2008.

[4] D. Song, N. Qin, and K. Goldberg, "A minimum variance calibration algorithm for pan-tilt robotic cameras in natural environments," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2006, pp. 3449–3456.

[5] K. H. Strobl and G. Hirzinger, "Optimal hand-eye calibration," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots*, 2006, pp. 4647–4653.

[6] J. Fraden, *Handbook of Modern Sensors: Physics, Designs, and Applications*. New York, NY, USA: Springer, 2004.

[7] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[8] H. Li, C. Chou, L. Fan, B. Li, D. Wang, and D. Song, "Robotic subsurface pipeline mapping with a ground-penetrating radar and a camera," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots*, Oct. 2018, pp. 3145–3150.

[9] H. Li, C. Chou, L. Fan, B. Li, D. Wang, and D. Song, "Toward automatic subsurface pipeline mapping by fusing a ground-penetrating radar and a camera," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 2, pp. 722–734, Apr. 2020.

[10] N. Muhammad and S. Lacroix, "Calibration of a rotating multi-beam lidar," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots*, 2010, pp. 5648–5653.

[11] D. J. Daniels, *Ground Penetrating Radar, 2nd Edition*, vol. 1. London, U.K.: IET, 2004.

[12] H. M. Jol, *Ground Penetrating Radar Theory and Applications*. Amsterdam, The Netherlands: Elsevier, 2008.

[13] A. Annan, *Electromagnetic Principles of Ground Penetrating Radar*, vol. 1. Amsterdam, The Netherlands: Elsevier, 2009.

[14] D. Goodman and S. Piro, *GPR Remote Sensing in Archaeology*, vol. 9, New York, NY, USA: Springer, 2013.

[15] T. Fukuda, Y. Hasegawa, K. Kosuge, K. Komoriya, F. Kitagawa, and T. Ikegami, "Environment-adaptive antipersonnel mine detection system-advanced mine sweeper," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots*, 2006, pp. 3618–3623.

[16] T. Evans, R. Ponticelli, E. Garcia, P. G. de Santos, and M. Armada, "A scanning robotic system for humanitarian de-mining activities," *Ind. Robot, Int. J.*, vol. 35, no. 2, pp. 133–142, 2008.

[17] R. S. Lim, H. M. La, and W. Sheng, "A robotic crack inspection and mapping system for bridge deck maintenance," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 2, pp. 367–378, Apr. 2014.

[18] H. M. La, N. Gucunski, S.-H. Kee, and L. Van Nguyen, "Data analysis and visualization for the bridge deck inspection and evaluation robotic system," *Vis. Eng.*, vol. 3, no. 1, 2015, Art. no. 6.

[19] P. Prasanna *et al.*, "Automated crack detection on concrete bridges," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 2, pp. 591–599, Apr. 2016.

[20] T. Le, S. Gibb, N. Pham, H. M. La, L. Falk, and T. Berendsen, "Autonomous robotic system using non-destructive evaluation methods for bridge deck inspection," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2017, pp. 3672–3677.

[21] P. Furgale, T. D. Barfoot, N. Ghafoor, K. Williams, and G. Osinski, "Field testing of an integrated surface/subsurface modeling technique for planetary exploration," *Int. J. Robot. Res.*, vol. 29, no. 12, pp. 1529–1549, Oct. 2010.

[22] T. Barfoot *et al.*, "Field testing of a rover guidance, navigation, and control architecture to support a ground-ice prospecting mission to mars," *Robot. Auton. Syst.*, vol. 59, no. 6, pp. 472–488, 2011.

[23] J. Civera, O. G. Grasa, A. J. Davison, and J. Montiel, "1-point RANSAC for EKF-based structure from motion," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2009, pp. 3498–3504.

[24] Y. Lu and D. Song, "Visual navigation using heterogeneous landmarks and unsupervised geometric constraints," *IEEE Trans. Robot.*, vol. 31, no. 3, pp. 736–749, Jun. 2015.

[25] R. Mur-Artal and J. D. Tardós, "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.

[26] F. M. Mirzaei, D. G. Kottas, and S. I. Roumeliotis, "3D lidar–camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization," *Int. J. Robot. Res.*, vol. 31, no. 4, pp. 452–467, 2012.

[27] F. Vasconcelos, J. P. Barreto, and U. Nunes, "A minimal solution for the extrinsic calibration of a camera and a laser-rangefinder," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2097–2107, Nov. 2012.

[28] C. Chou, S. Yeh, J. Yi, and D. Song, "Extrinsic calibration of a ground penetrating radar," in *Proc. IEEE Int. Conf. Autom. Sci. Eng.*, 2016, pp. 1326–1331.

[29] C. Chou, S. Yeh, and D. Song, "Mirror-assisted calibration of a multi-modal sensing array with a ground penetrating radar and a camera," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots*, Sep. 2017, pp. 1457–1463.

[30] C. Chou, A. Kingery, D. Wang, H. Li, and D. Song, "Encoder-camera-ground penetrating radar tri-sensor mapping for surface and subsurface transportation infrastructure inspection," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2018, pp. 1452–1457.

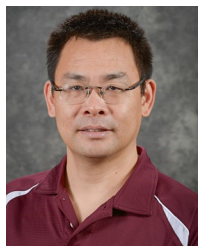
[31] GSSI, *Concrete Handbook (MN72-367 Rev H)*. Nashua, NH, USA: Geophys. Surv. Syst., Inc., 2017.
[32] C. Warren and A. Giannopoulos, "Investigation of the directivity of a commercial ground-penetrating radar antenna using a finite-difference time-domain antenna model," in *Proc. 14th Int. Conf. Ground Penetrating Radar*, 2012, pp. 226–231.
[33] N. Diamanti and A. P. Annan, "Air-launched and ground-coupled GPR data," in *Proc. IEEE 11th Eur. Conf. Antennas Propag.*, 2017, pp. 1694–1698.
[34] N. Diamanti, A. P. Annan, and J. D. Redman, "Concrete bridge deck deterioration assessment using ground penetrating radar (GPR)," *J. Environ. Eng. Geophys.*, vol. 22, no. 2, pp. 121–132, 2017.
[35] G. S. Baker, T. E. Jordan, and J. Pardy, "An introduction to ground penetrating radar (GPR)," *Special Paper Geolog. Soc. Amer.*, vol. 432, pp. 1–18, 2007.
[36] A. Neal, "Ground-penetrating radar and its use in sedimentology: principles, problems and progress," *Earth-Sci. Rev.*, vol. 66, no. 3/4, pp. 261–330, 2004.
[37] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An accurate $O(n)$ solution to the PnP problem," *Int. J. Comput. Vision*, vol. 81, no. 2, pp. 155–166, 2008.
[38] S. Li, C. Xu, and M. Xie, "A robust $O(n)$ solution to the perspective-n-point problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1444–1450, Jul. 2012.
[39] P. Lébraly, C. Deymier, O. Ait-Aider, E. Royer, and M. Dhome, "Flexible extrinsic calibration of non-overlapping cameras using a planar mirror: Application to vision-based robotics," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots*, 2010, pp. 5640–5647.
[40] R. Rodrigues, J. P. Barreto, and U. Nunes, "Camera pose estimation using images of planar mirror reflections," in *Proc. Eur. Conf. Comput. Vision*, 2010, pp. 382–395.
[41] A. Agha-Mohammadi, and D. Song, "Robust recognition of planar mirrored walls using a single view," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 1186–1191.
[42] Y. Lu, D. Song, H. Li, and J. Liu, "Automatic recognition of spurious surface in building exterior survey," in *Proc. IEEE Int. Conf. Autom. Sci. Eng.*, Aug. 2013, pp. 1047–1052.
[43] G. Long, L. Kneip, X. Li, X. Zhang, and Q. Yu, "Simplified mirror-based camera pose computation via rotation averaging," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, 2015, pp. 1247–1255.
[44] R. Hartley, J. Trumpf, Y. Dai, and H. Li, "Rotation averaging," *Int. J. Comput. Vision*, vol. 103, no. 3, pp. 267–305, 2013.
[45] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[46] B. Horn, "Closed form solution of absolute orientation using unit quaternions," *J. Opt. Soc. A*, vol. 4, no. 4, pp. 629–642, 1987.
[47] J.-Y. Bouguet, "Camera calibration toolbox for MATLAB," 2004. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc
[48] R. Kümmerle *et al.*, "On measuring the accuracy of SLAM algorithms," *Auton. Robots*, vol. 27, no. 4, 2009, Art. no. 387.
[49] W. Burgard *et al.*, "A comparison of SLAM algorithms based on a graph of relations," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2009, pp. 2089–2095.

Chieh Chou received the B.S. and M.S. degrees in electronic engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2008 and 2010, respectively, and the Ph.D. degree in computer science and engineering from Texas A&M University, College Station, TX, USA, in 2019.

His research interests include computer vision, sensor fusion, robot perception, and localization.

Haifeng Li (Member, IEEE) received the Ph.D. degree in control theory and control engineering from Nankai University, Tianjin, China, in 2012.

He is currently an Associate Professor with the Department of Computer Science and Technology, Civil Aviation University of China, Tianjin. He has authored or coauthored more than 30 technical articles. His research interests include computer vision, image processing, robotic sensing, multisensor fusion, robot localization, and navigation.



Dezhen Song (Senior Member, IEEE) received the Ph.D. degree in industrial engineering from the University of California, Berkeley, Berkeley, CA, USA, in 2004.

He is currently a Professor with the Department of Computer Science and Engineering, Texas A&M University, College Station, TX, USA. His research interests include networked robotics, distributed sensing, computer vision, surveillance, and stochastic modeling.

Dr. Song was an Associate Editor for the IEEE TRANSACTIONS ON ROBOTICS, from 2008 to 2012. From 2010 to 2014, he was an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING. He is currently a Senior Editor for the IEEE ROBOTICS AND AUTOMATION LETTERS. He was the recipient of the Kayamori Best Paper Award of the 2005 IEEE International Conference on Robotics and Automation (with J. Yi and S. Ding). He was also the recipient of the NSF Faculty Early Career Development (CAREER) Award in 2007.