

Accuracy of inside-out visual-inertial tracking pipelines for freehand 3D ultrasound probe localization

Zohar Oddes¹ and Dana Solav¹

¹Faculty of Mechanical Engineering, Technion - Israel Institute of Technology, Haifa, Israel

ABSTRACT

Freehand three-dimensional ultrasound often requires accurate probe localization. Inside-out visual-inertial tracking mounts the camera and inertial unit on the probe itself, offering a compact alternative to external systems. Achievable accuracy depends on the complete pipeline, not the sensor alone, and comparisons rarely state which differences their setup can resolve. We compared three pipelines under identical conditions: the ZED X Mini with the ZED SDK, the ZED X Mini with cuVSLAM, and the Intel RealSense D455 with cuVSLAM. Both cameras and an infrared-reflective marker cluster were mounted on a rigid frame representing the probe base, tracked by an eleven-camera motion-capture system. Four recordings covered three motion conditions of decreasing constraint, from pivoting about a fixed point to full freehand manipulation, one repeated to quantify trial-to-trial variability. All processing was offline. Trajectories were synchronized and aligned by robot-world/hand-eye calibration, and accuracy assessed from per-frame residuals. The ZED X Mini with cuVSLAM gave the lowest median rotational residual in every condition, the lowest median translational residual in most (2.5–4.1 mm and 0.100–0.117°), and the lowest 95th-percentile residual in every condition and component. Part of that advantage reflects cuVSLAM returning a full-sequence optimized trajectory, whereas the ZED SDK returns causal per-frame poses. Differences between motion conditions largely fell within the resolution of the comparison, derived from reference uncertainty and the precision of each statistic, the exception being a higher translational residual for ZED X Mini with cuVSLAM under freehand motion. Pipeline choice, more than motion condition, is therefore the dominant determinant of tracking accuracy, and the ZED X Mini with cuVSLAM is the most suitable of the three. For freehand 3D-ultrasound applications, the rotational error is small relative to reported probe-to-image-plane calibration error, whereas the millimeter-level translational error dominates overall pose accuracy. Whether either is sufficient for volumetric reconstruction remains to be established.

Keywords: freehand 3D ultrasound, hand-eye calibration, motion capture, probe tracking, SLAM, visual-inertial odometry

1 INTRODUCTION

Three-dimensional anatomical reconstruction from freehand ultrasound has become an important tool for image-guided diagnosis and navigation, surgical planning, and quantitative musculoskeletal assessment (Tümer et al., 2018; Lee, 2014; Gebhardt et al., 2023; Weide et al., 2017; Habersack et al., 2022). By sweeping a 2D ultrasound probe across anatomy while tracking its spatial pose, successive B-mode frames can be compounded into volumetric representations of patient-specific structures (Mozaffari and Lee, 2017; Peng et al., 2022). Placing a frame within the volume requires two transformations: the pose of the probe in space, and the fixed transformation from the probe to its image plane, the latter obtained by a separate spatial calibration for which an established literature and open-source tooling exist (Mercier et al., 2005; Lasso et al., 2014). However, the quality and spatial consistency of the resulting volumes are strongly dependent on the accuracy of probe localization during acquisition.

A further application, and the motivation for the present work, is patient-specific biomechanical modeling. Finite-element models of the residual limb used to inform prosthetic socket design have been constructed from magnetic-resonance-derived geometry combined with in-vivo indentation (Sengeh et al., 2016; Moerman et al., 2019). Magnetic resonance imaging is, however, costly and largely unavailable in the clinics where such devices are fitted, and ultrasound has accordingly been proposed as a portable alternative for acquiring the same information (Ranger et al., 2019; Fougerson et al., 2020). It has since been used to supply tissue material properties (Ranger et al., 2023) and, with the probe itself acting as the indenter, the internal

tissue displacements that drive inverse identification (Fougeron et al., 2020; Doherty et al., 2022; Schimmoeller et al., 2019, 2020). Where the probe is manipulated freehand, its pose must then be recovered: Doherty et al. (2022) track an instrumented probe carrying a load cell, an inertial measurement unit and an optical marker cluster (Schimmoeller et al., 2019, 2020), and reduce the measured motion to a straight-line displacement between its start and end positions, assuming no rotation of the probe. The parameters recovered by such procedures are sensitive to the measurements that constrain them (Oddes and Solav, 2023; Fougeron et al., 2024), and because the ultrasound observable is sampled in the image plane of the probe, the placement of that plane relative to the tissue enters the identification directly. Probe positioning error has accordingly been identified as a contributor to model-fidelity error in this setting, but one that remains unquantified (Fougeron et al., 2024). Establishing how accurately a freehand probe can be localized is therefore a prerequisite for extending these methods beyond constrained, single-site indentation.

Traditional freehand ultrasound systems rely on external tracking devices such as optical motion capture, electromagnetic tracking, or mechanical articulated arms to estimate probe pose (Maybody et al., 2013; Fattori et al., 2021). While these systems can achieve sub-millimeter accuracy (Koivukangas et al., 2013), they are often expensive, require careful setup (Wang et al., 2013), and may impose constraints such as line-of-sight requirements or sensitivity to magnetic interference. In contrast, inside-out tracking approaches integrate sensors such as stereo cameras and inertial measurement units (IMUs) directly onto the probe, enabling self-localization through visual-inertial odometry (VIO) and simultaneous localization and mapping (SLAM). These systems offer improved portability and ease of deployment, making them increasingly attractive for clinical and point-of-care applications (Sankepalle et al., 2023).

Inside-out systems differ, first, in what they localize against. Probe-mounted cameras have been used to track features of the patient’s own skin (Sun et al., 2014) or the surrounding external anatomy (Wang et al., 2017), which renders the estimated pose intrinsically robust to rigid motion of the patient but restricts the observed scene to a small surface patch. A third option prepares the environment rather than relying on the patient or on the room as found: Zhang et al. (2025) localize two probe-mounted cameras against a printed textured workspace, which yields an absolute pose in a global frame and so accumulates no drift along a sweep, but requires that workspace to be installed and confines the sweep to the area it covers. Visual-inertial pipelines of the kind evaluated in the current study, instead localize against the surrounding room, which places no restriction on probe motion but refers the trajectory to a frame the patient may move within.

A second distinction concerns the scale at which such pipelines are characterized. Off-the-shelf visual-inertial systems, including stereo cameras of the type used here, have most often been evaluated on mobile robots or vehicles, where reported errors are of the order of decimeters over trajectories of roughly a dozen meters (Alapetite et al., 2020). Freehand ultrasound occupies a different regime: the anatomical region swept is typically limb- or organ-scale, centimeters to decimeters across, but the cumulative path traced by the probe over a complete acquisition can still amount to several meters, while the relevant error scale is millimetric. Several probe-tracking studies instead report error as a function of distance traveled — either a per-distance drift rate, or an error accumulated over just the first few centimeters of motion (Sun et al., 2014; Wang et al., 2017) — characterized over a small fraction of the distance a complete sweep actually covers, so they do not establish how error behaves once travel extends to that scale. Accuracy figures obtained at one scale transfer poorly to another, and a comparison intended for freehand ultrasound must be conducted at the scale of the intended use.

A wide range of probe tracking systems and freehand ultrasound frameworks have been proposed in recent years, spanning optical, inertial, and hybrid sensing approaches (Peng et al., 2022; Adriaans et al., 2024). However, evaluating and comparing these systems remains challenging. First, performance depends not only on sensor specifications, but also on the full processing pipeline, including acquisition hardware, synchronization strategy, and SLAM algorithm configuration. Second, reported evaluations are often performed under a single motion scenario, so it is not known whether a ranking obtained under one pattern of probe motion transfers to another. Tracking performance may depend on the characteristics of probe motion, including the degree of constraint and the presence of rapid or irregular changes, and evaluations rarely vary these characteristics in a controlled way. Whether pipeline differences persist across motion regimes, and how large a motion effect would have to be before it could be distinguished from measurement uncertainty, therefore remain open. A third difficulty is that the reference against which accuracy is measured is itself seldom characterized (Stolkin et al., 2006). Reported evaluations commonly adopt an optical tracker, a robotic arm or a motion stage as ground truth and describe its accuracy by a manufacturer specification, or by an unquantified assertion of sub-millimeter performance (Sun et al., 2014; Wang et al., 2017; Benjamin et al., 2020; He et al., 2024; Zhang et al., 2025). The uncertainty contributed by the reference itself, by the registration of the estimated trajectory to it, and by their temporal alignment is rarely propagated into the reported errors, although each enters them directly. This matters most where it is least visible: when the quantity being measured approaches the accuracy of the instrument measuring it, small differences between systems cease to be interpretable. Establishing which differences a given experimental arrangement can

actually resolve is therefore a prerequisite for comparing pipelines at the millimeter and sub-degree level.

To address these challenges, this study evaluates complete visual-inertial pose-tracking pipelines under structured motion conditions using a motion-capture-based reference system. The experimental protocol included three motion conditions with different levels of constraint, ranging from a constrained pivot motion to unconstrained freehand manipulation and a mixed condition combining the two, with one condition acquired twice to quantify trial-to-trial repeatability. The motion conditions were designed to provide complementary and repeatable variations in probe motion, spanning the range over which a probe-mounted tracker would be used, rather than to constitute a standardized model of clinical ultrasound scanning. We compare three state-of-the-art inside-out tracking pipelines that combine stereo vision and inertial sensing with different SLAM implementations. Rather than isolating individual components, we focus on full end-to-end tracking pipelines, including hardware, software, and processing stack. The study focuses on the accuracy and consistency of probe pose estimation and does not directly evaluate the effect of tracking errors on downstream ultrasound image reconstruction.

The contributions of this study are as follows:

- A systematic comparison of complete visual-inertial tracking pipelines for freehand probe localization.
- An assessment of whether the relative performance of the pipelines is stable across motion conditions ranging from constrained pivoting to unconstrained freehand manipulation.
- An analysis of translational and rotational tracking accuracy relative to a motion capture reference, including an explicit derivation of the reference uncertainty and of the smallest differences the comparison can resolve.
- A repeatability measurement from a duplicated acquisition, bounding trial-to-trial variation in the reported errors.
- A comparison of practical pipeline characteristics, including pose output rate, pose-stream continuity, and the reliability of temporal alignment to an external reference.

2 METHODS

2.1 Tracking pipelines

The study was designed to compare complete visual-inertial pose-tracking pipelines rather than isolated camera hardware. A pose-tracking pipeline was defined as the combination of (i) the camera and its integrated IMU, (ii) the acquisition platform, and (iii) the visual-inertial localization algorithm used to estimate the six-degree-of-freedom trajectory. The three pipelines considered in this study are summarized in [Table 1](#) and detailed in [section 2.5](#).

Table 1. Pose tracking pipelines evaluated in this study.

Pipeline	Camera	Acquisition platform	Localization framework
I. ZED-SDK	ZED X Mini	Jetson Orin NX	ZED SDK Positional Tracking
II. ZED-cuVSLAM	ZED X Mini	Jetson Orin NX	cuVSLAM (Isaac ROS)
III. RS-cuVSLAM	D455	Workstation	cuVSLAM (Isaac ROS)

2.2 Experimental setup and reference tracking

A custom rigid frame representing the ultrasound-probe base was constructed to allow the two tracking cameras to be mounted as a single rigid assembly, as shown in [Figure 1a](#). The ZED X Mini and Intel RealSense D455 cameras were mounted on the frame such that their relative poses with respect to the probe base remained fixed throughout each trial. A set of infrared-reflective markers was also rigidly attached to the frame to enable external motion tracking using a Vicon motion-capture system (Vicon Motion Systems Ltd., Oxford, UK) consisting of 11 cameras operating at 240 Hz. The reflective markers were arranged as a rigid cluster on the experimental frame, and the three-dimensional position of the cluster was reconstructed using Vicon Nexus v2.16. The resulting rigid-body trajectory was used as the external reference for evaluating the visual-inertial tracking pipelines. The reference trajectory represents the motion of the rigid frame rather than the optical center of either tracking camera. The spatial relationship between the Vicon marker cluster and each camera was therefore estimated during the calibration procedure described in [section 2.6](#). The resulting rigid transformation was used to express the camera trajectory in the Vicon coordinate frame of each tracking pipeline before calculating trajectory errors.

2.3 Test protocol

Four recordings were made, covering three motion conditions, in which the frame was manually manipulated within the Vicron capture volume while the scene depicted in the background of [Figure 1b](#) remained mostly in view of the cameras. The pivot condition was recorded twice, providing a repeatability measurement under a nominally identical protocol. The tripod mounting assembly depicted in the foreground of [Figure 1b](#) consists of stationary tripod legs and a 60 cm long rigid extension bar connected through a ball-and-socket adapter. For the pivot test, the extension bar was attached to the frame, constraining the motion to pivot about a fixed point in space at a distance of approximately 160 cm from the wall. For the freehand test, the extension bar was detached from the tripod legs and the frame was manipulated using unconstrained freehand sweeping movements at a similar distance from the wall. In the mixed test, the camera frame movement transitioned from a pivot motion to freehand motion by detaching the extension bar mid-test. To aid synchronization between Vicron tracking data and visual-inertial tracking data in post-processing, all tests began and ended by holding the extension bar upright and panning the cameras back and forth around the vertical axis ([Schmedling et al., 2026](#)).

The resulting probe motions are characterized quantitatively in [Table 2](#). All quantities were computed from the Vicron marker-cluster trajectory within the shared evaluation window of each test, and are therefore independent of any tracking pipeline. The three conditions were acquired at comparable angular rates, while the translational path length increased from 9.7 m under the constrained pivot motion to 12.9 and 13.4 m as the constraint was released. The constrained pivot did not, however, produce the least rotation. Its angular extent of 108° exceeds the 77° of the freehand recording, and its median angular rate was the highest of the three. The conditions therefore differ in their translational rather than their rotational dynamics. The same pattern holds throughout. As the constraint was released the median linear speed rose from 168 to 269 mm/s and the median linear acceleration from 301 to 697 mm/s², so that the ratio of the two, a characteristic frequency of the motion, separates the two pivot recordings (1.80 and 1.71 s⁻¹) from the mixed and freehand recordings (2.61 and 2.59 s⁻¹). The unconstrained motions were thus not only faster but less smooth. Accelerations are reported for comparison between conditions rather than as absolute values, since they depend on the differentiation window (as opposed to the linear speeds, which vary by less than 1% over windows from 0.05 to 0.40 s). The repeated pivot recording has a shorter shared evaluation window (37.4 s) owing to an interruption in one of the acquisitions, and differs from the first pivot recording in both angular extent and calibration margin, so the two pivot trials are repetitions of the protocol rather than of an identical motion.

Table 2. Characteristics of the probe motion in each recording, computed from the Vicron marker-cluster trajectory within the shared evaluation window and therefore independent of any tracking pipeline. The final row is a second acquisition of the pivot condition, used to assess repeatability, and is not an additional motion condition. The angular extent is the largest angle between any two orientations occurring within the window; ω and $\|\mathbf{v}\|$ denote the angular rate and the linear speed of the marker cluster, the latter obtained by Savitzky–Golay differentiation ([Savitzky and Golay, 1964](#); [Virtanen et al., 2020](#)) (third order, 0.10 s window); and $1 - \sigma_2$ is the margin by which the hand-eye calibration is determined, defined in [section 2.6](#), with zero indicating a degenerate motion.

Recording	Duration [s]	Angular extent [deg]	ω [deg/s]		$\ \mathbf{v}\ $ [mm/s]		Path [m]	Calibration margin $1 - \sigma_2$
			Median	95th	Median	95th		
Pivot	55.1	108.3	25.0	58.6	168	428	9.74	0.031
Freehand	46.9	77.3	21.0	56.8	269	538	13.39	0.029
Mixed	59.7	98.1	21.7	62.1	209	525	12.88	0.036
Pivot (repeat)	37.4	116.1	23.6	51.7	201	386	6.97	0.044

2.4 Data acquisition

ZED X Mini The 2.2 mm lens ZED X Mini stereo camera (Stereolabs, San Francisco, CA, USA) was connected to an NVIDIA Jetson Orin Nano Super Developer Kit fitted with a Jetson Orin NX 16 GB module (NVIDIA, Santa Clara, CA, USA). ZED Software Development Kit (ZED SDK v5.0.7) operating on Ubuntu 22.04 LTS via JetPack 6.2.1 was used to record 1920×1200 pixel-resolution stereo images at 60 Hz and IMU data at 400 Hz, which were stored in the native SVO2 format on a local NVMe SSD. The SVO2 files were subsequently transferred to the workstation for offline processing and pose estimation.

RealSense D455 The D455 stereo camera (Intel, Santa Clara, CA, USA) with firmware version 5.13.0.50 was connected directly to the workstation via a USB 3.2 cable. The RealSense Viewer application (v2.56.5) was used to record native RealSense-bag files containing infrared stereo images with 1280×800 pixel-resolution (8-bit monochrome) and IMU data at a rate of 30 Hz and 400 Hz, respectively. For subsequent processing with the cuVSLAM pipelines presented below, the

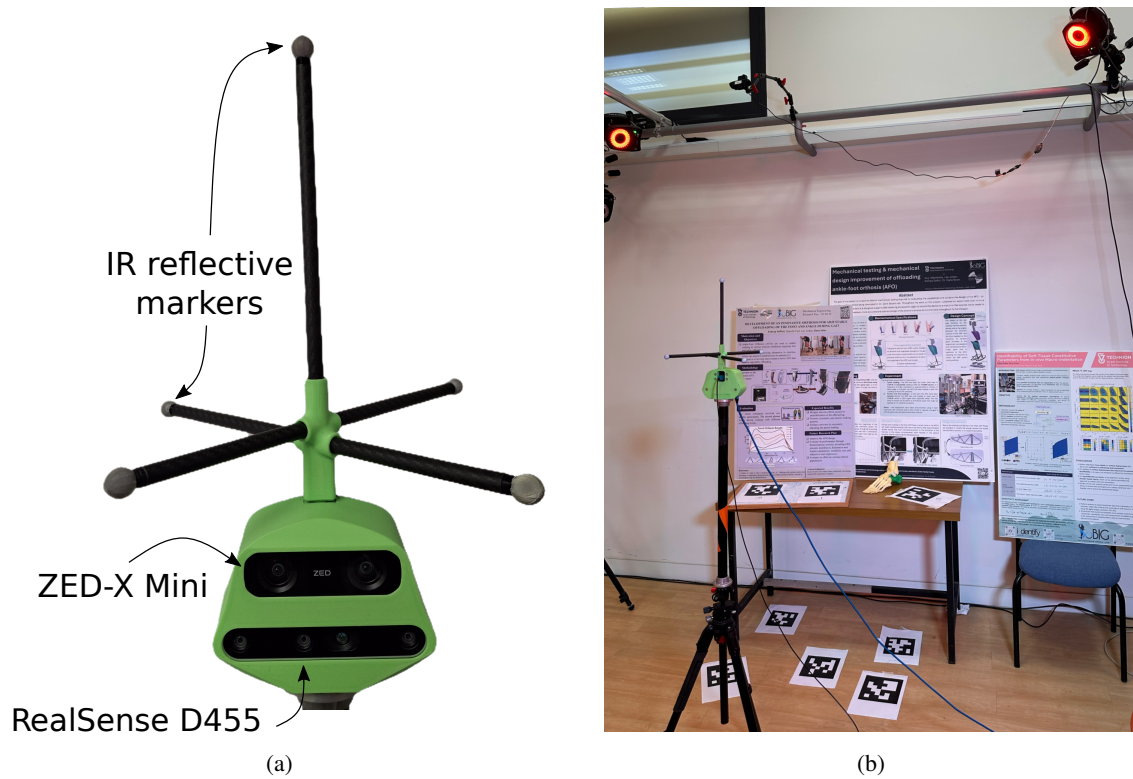


Figure 1. Experimental setup. (a) Custom frame representing the ultrasound-probe base; both stereo cameras and the infrared-reflective markers are mounted to the frame so that they move as a single rigid body. (b) Test environment, showing the static high-contrast scene used to support visual odometry (background) and the mounting assembly used to constrain the probe motion (foreground), consisting from top to bottom of the custom frame (green), a rigid extension bar, a ball-and-socket joint, and stationary tripod legs.

RealSense-bag files were directly converted to ROS 2-bag files by a custom Python script (see [Code availability](#)).

2.5 Visual-inertial pose estimation

Each pipeline produced a time-stamped list of camera poses which was stored in a common CSV format for consistent handling across pipelines. For every processed frame, the CSV file contained the frame index, a timestamp inherited from the acquisition platform clock, the translation and rotation vectors, as well as other pipeline-specific diagnostic fields where available, such as tracking state and pose-confidence metrics. Offline processing for all pipelines was performed on the same Linux workstation (Ubuntu 22.04.5 LTS, 64-bit, 32.0 GiB RAM, 11th Gen Intel® Core™ i9-11900K, NVIDIA GeForce RTX 3050).

ZED SDK For the ZED-SDK pipeline, SVO2 recordings from the ZED X Mini were decoded and processed using the positional-tracking module of the ZED SDK v5.3.0, accessed through the ZED Python API (pyzed v5.3) and a custom script. For each SVO2 file, the script initialized the camera in offline playback mode and enabled visual-inertial SLAM with `POSITIONAL_TRACKING_MODE.GEN_3`, IMU fusion enabled (`enable_imu_fusion = true`) and spatial memory enabled (`enable_area_memory = true`). All remaining tracking parameters were left at their default values. The script iterated over each recorded frame and, whenever the tracking state was valid, wrote a row to a CSV file in the common format described above. Poses were thus recorded as the tracking module produced them. The ZED SDK Python API provides no means of retrieving a globally optimized trajectory once playback has completed, so unlike the cuVSLAM trajectories these poses are not refined using later observations.

cuVSLAM For both ZED-cuVSLAM and RS-cuVSLAM pipelines, processing was done with NVIDIA Isaac ROS release 3.2 inside a Docker container based on the Isaac ROS Development Environment image (`isaac_ros_dev`) modified to support both ZED and RealSense frameworks. The ZED and RealSense ROS 2 wrappers were used to play back the recorded data from SVO2 and ROS 2 bag files, respectively, and to publish stereo image and IMU topics with their original timestamps. An Isaac ROS visual SLAM node implementing cuVSLAM (Korovko et al., 2025) subscribed to these topics and estimated

the camera pose using the default stereo-inertial settings. When playback reached the end of each file, the optimized camera trajectory maintained by cuVSLAM was exported to a CSV file in the common format described above. To reduce frame loss in the visual SLAM node due to processing load, all recordings were played back at 40% of their original frame rate while preserving the original acquisition timestamps via ROS 2 simulated time (`use_sim_time = true`). ROS 2 queue and buffer sizes were increased to provide a sufficient buffer for the incoming stereo and IMU streams.

2.6 Coordinate frames, trajectory alignment and error metrics

Rigid-body transformations are denoted as ${}^A\mathbf{T}_B$, representing the pose of frame B expressed in frame A . Each transformation is composed of a rotation matrix, ${}^A\mathbf{R}_B \in SO(3)$, and a translation vector, ${}^A\mathbf{t}_B \in \mathbb{R}^3$, such that

$${}^A\mathbf{T}_B = \begin{bmatrix} {}^A\mathbf{R}_B & {}^A\mathbf{t}_B \\ \mathbf{0}^\top & 1 \end{bmatrix}.$$

The analysis uses four coordinate frames, related as shown in Figure 2. The Vicon frame V is the fixed laboratory reference defined by the motion-capture system. The marker frame M is that of the infrared-reflective marker cluster, whose pose ${}^V\mathbf{T}_M(t)$ is derived using the Kabsch algorithm. The camera frame C is the frame whose pose ${}^W\mathbf{T}_C(t)$ each pipeline reports: for ZED-SDK the left optical sensor in the ZED SDK’s right-handed Y-up convention, and for the cuVSLAM pipelines the configured base frame (`base_link` for ZED-cuVSLAM, `camera_link` for RS-cuVSLAM) in the ROS convention. These differ by constant rigid transformations, which the calibration below absorbs and which therefore do not affect the reported residuals. The SLAM world frame W is initialized independently for each recording, at the camera pose when tracking begins.

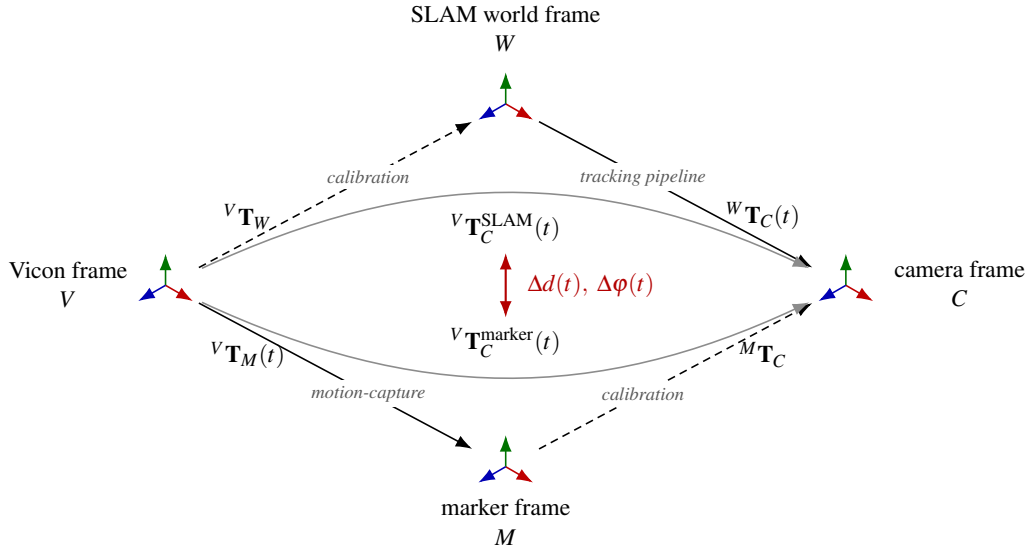


Figure 2. Two paths from the Vicon frame V to the camera frame C , obtained by the robot-world/hand-eye calibration. The upper path represents the SLAM-based camera pose estimate (Equation 1) and the lower path the marker-based camera pose estimate (Equation 2). The difference between the two branches is defined as the calibration residual, from which the translational error (Equation 3) and the rotational error (Equation 5) are derived.

The tracking cameras and the Vicon system run on independent time bases and sampling rates, so the trajectories were first temporally synchronized. Within each test, the per-pipeline temporal offset was estimated from the angular-velocity magnitudes of the full Vicon marker trajectory ${}^V\mathbf{T}_M(t)$ and the corresponding SLAM trajectory ${}^W\mathbf{T}_C(t)$ (Cao et al., 2026). The Vicon trajectory was then interpolated onto the camera timestamps after applying the estimated shift. The angular-velocity magnitude is the only criterion available before the spatial calibration exists, and being a scalar, it constrains the offset less tightly than the pose trajectories themselves. The offset was therefore refined after calibration by minimizing the median rotational residual and recalibrating at the refined value, which changed the estimate by less than 3 ms for the cuVSLAM pipelines but by up to 11.8 ms for ZED-SDK, whose scalar objective is the shallowest and localizes the offset least tightly. The full synchronization scheme is detailed in Appendix A.

The two trajectories are expressed in different frames and cannot be compared directly: ${}^W\mathbf{T}_C(t)$ gives the camera pose in the SLAM world frame, ${}^V\mathbf{T}_M(t)$ the marker pose in the Vicon frame. Bringing both into the Vicon frame requires two further

transformations, ${}^V\mathbf{T}_W$, the pose of the SLAM world frame in the Vicon frame, and ${}^M\mathbf{T}_C$, the pose of the camera frame relative to the marker cluster. Neither is measured, but both are constant within a recording: W is fixed once tracking begins, and the tracking camera and marker cluster were rigidly attached throughout each test. That constancy is what makes them recoverable from the trajectories themselves, since at every frame

$${}^V\mathbf{T}_W {}^W\mathbf{T}_C(t) = {}^V\mathbf{T}_M(t) {}^M\mathbf{T}_C,$$

both sides expressing the same camera pose in the Vicon frame. Each frame therefore constrains the same twelve unknowns, and a recording of many frames determines them. This is the robot-world/hand-eye calibration problem, solved here using OpenCV's `calibrateRobotWorldHandEye` function with the `CALIB_ROBOT_WORLD_HAND_EYE_SHAH` method (Bradski, 2000; Shah, 2013).

Such a calibration is well posed only if the motion rotates about more than one axis. Because the problem was solved by Shah's method, this requirement can be evaluated directly on the matrix that method factorizes. Its largest singular value is unity by construction and is attained at the calibration itself, so a second unit singular value would indicate a second consistent solution and hence a degenerate motion. The margin $1 - \sigma_2$ therefore measures how well the calibration is determined, and it depends only on the reference trajectory, which makes it identical for all three pipelines within a recording. Appendix B gives the derivation. Across the four recordings the margin ranged from 0.029 to 0.044 (Table 2), so every motion was well determined. The variation between motion conditions (0.029–0.036) was moreover smaller than that between the two repetitions of the pivot condition (0.031 and 0.044), so differences in calibration conditioning do not account for the differences in tracking error reported below.

After calibration, the SLAM-based camera pose can be expressed in the Vicon frame as

$${}^V\mathbf{T}_C^{\text{SLAM}}(t) = {}^V\mathbf{T}_W {}^W\mathbf{T}_C(t), \quad (1)$$

while the marker-based estimate of the camera pose is given by

$${}^V\mathbf{T}_C^{\text{marker}}(t) = {}^V\mathbf{T}_M(t) {}^M\mathbf{T}_C. \quad (2)$$

Their disagreement defines the calibration residual. The translational error is the Euclidean distance between the translation components,

$$\Delta d(t) = \left\| {}^V\mathbf{t}_C^{\text{SLAM}}(t) - {}^V\mathbf{t}_C^{\text{marker}}(t) \right\|. \quad (3)$$

The rotational error follows from the relative rotation between the two poses,

$$\mathbf{R}_{\text{err}}(t) = ({}^V\mathbf{R}_C^{\text{marker}}(t))^T ({}^V\mathbf{R}_C^{\text{SLAM}}(t)), \quad (4)$$

as its angle of rotation

$$\Delta\phi(t) = \arccos\left(\frac{\text{tr}(\mathbf{R}_{\text{err}}(t)) - 1}{2}\right). \quad (5)$$

Because ${}^V\mathbf{T}_W$ and ${}^M\mathbf{T}_C$ are estimated from the same paired trajectories, the marker-based trajectory is not an independent ground truth: $\Delta d(t)$ and $\Delta\phi(t)$ measure the disagreement remaining after aligning the SLAM trajectory to the Vicon-measured marker motion. Calibration and error evaluation were restricted throughout to the shared temporal window in which valid SLAM data were available for all three pipelines.

Successive frames sample almost the same tracking state, so the residual series are strongly autocorrelated and the frame count overstates the information they carry. The integrated autocorrelation time and the resulting effective sample size are given in Appendix A. Confidence intervals for the medians and 95th percentiles reported below were therefore obtained by a moving-block bootstrap (Künsch, 1989), rather than from the effective sample size, whose value depends strongly on how the correlation time is defined (Theunissen et al., 2008). Each replicate was assembled by drawing blocks of consecutive frames, each spanning three correlation times, from random starting positions and with replacement, and concatenating them until the replicate matched the length of the original series. The statistic was then recomputed on it. Resampling in blocks rather than frame by frame preserves the dependence within each block, without which the interval would assume as many independent observations as there are frames. The intervals quoted in Table 4 span 1.96 standard deviations of the statistic over 2×10^4 such replicates, a normal-approximation interval shown to give coverage comparable to the percentile interval for correlated

inertial-sensor data (Sabatini, 2006). To test for a difference between two cells, that difference was itself resampled. For two pipelines on the same recording, a single set of block start times was drawn per replicate and applied to both series, which covered the same window even though they were sampled at different rates, so that any variation the two shared would cancel in the difference rather than be assumed away. For the same pipeline on different recordings, there was no common stretch of motion to pair, and the blocks were drawn independently. In both cases a difference is treated as established only where the resulting 95% interval for the difference excludes zero. Consequently, a difference between two cells may still be established despite their overlapping confidence intervals. Testing for overlap is more conservative than testing the difference itself, and more so when the two estimates are positively correlated (Schenker and Gentleman, 2001), which is why the difference was resampled directly rather than inferred from the two intervals.

3 RESULTS

Before evaluating the SLAM-based trajectories, the quality of the Vicon marker-cluster tracking was assessed from the residuals of the rigid-body fit. For each Vicon frame, the RMS residual of the Kabsch rigid-body fit across the five markers was computed, and summary statistics of these per-frame RMS values were calculated within the shared evaluation window of each test, as summarized in Table 3. Across all recordings, the mean RMS residual remained below 0.28 mm and the maximum RMS residual below 0.74 mm, indicating consistent marker-cluster tracking throughout.

Table 3. Summary statistics of the per-frame root mean square (RMS) residual of the Vicon rigid-body fit within the shared evaluation window of each recording. The final row is a second acquisition of the pivot condition, used to assess repeatability, and is not an additional motion condition.

Recording	Duration [s]	RMS residuals [mm]	
		Mean $\pm$ SD	Maximum
Pivot	55.1	0.279 ± 0.088	0.518
Freehand	46.9	0.231 ± 0.066	0.733
Mixed	59.7	0.209 ± 0.058	0.414
Pivot (repeat)	37.4	0.265 ± 0.067	0.465

The SLAM-based tracking accuracy was evaluated across three probe-motion conditions: pivot, freehand, and combined pivot-freehand motion. For each pipeline, the translational and rotational hand-eye calibration residuals were computed within the shared evaluation window of each test, as defined in Equations 3 and 5. To characterize the distribution and variability of calibration accuracy across the three pipelines, we analyzed the error distributions for each probe-motion type using violin plots, which simultaneously display the full distribution and quartiles. Figure 3 presents the calibration residuals stratified by pipeline and probe-motion condition, with translational errors shown on the left y-axis (millimeters) and rotational errors on the right y-axis (degrees). The whiskers of the box plots extend up to 1.5 times the interquartile range (IQR), which covers nearly the entire data range.

All residual statistics are summarized in Table 4. In the Pivot and Mixed conditions, ZED-cuVSLAM had the lowest median translational error (2.63 and 2.51 mm), while RS-cuVSLAM was marginally lower in the Freehand condition (3.81 mm against 4.09 mm). In both the Pivot and Freehand conditions, ZED-SDK produced the highest (5.26 and 4.77 mm). ZED-cuVSLAM had the lowest median rotational error in all three conditions (0.100° , 0.115° and 0.117°). It also had the lowest 95th percentile in every condition and both components. The ordering of the two remaining pipelines varied with condition and component.

Because the camera-to-marker transformation ${}^M\mathbf{T}_C$ describes a fixed physical geometry, it should be identical in every recording, and its variation between them measures the error of the calibration itself. Expressed in the marker-intrinsic frame, so that the arbitrary reference orientation of each rigid-body fit cancels, the transforms fitted in different motion conditions differ by 5.0–10.0 mm and 0.12 – 0.26° , with ZED-SDK the most variable in translation and RS-cuVSLAM the most variable in orientation. That variation overstates the influence of the calibration on the reported residuals, because ${}^M\mathbf{T}_C$ and ${}^V\mathbf{T}_W$ are fitted together and are strongly coupled (i.e., a displaced lever arm can be partly offset by a displaced world origin). Substituting the transform fitted in one recording for that of another, and re-solving only ${}^V\mathbf{T}_W$, which cannot be transferred because the SLAM world frame is initialized afresh in each recording, raises the median translational residual by 0.17 to 0.32 mm on average and by at most 1.05 mm. The calibration therefore contributes a few tenths of a millimeter to a reported median, an order of magnitude less than the spread of the transform that produces it.

Translational and Rotational Error Distributions

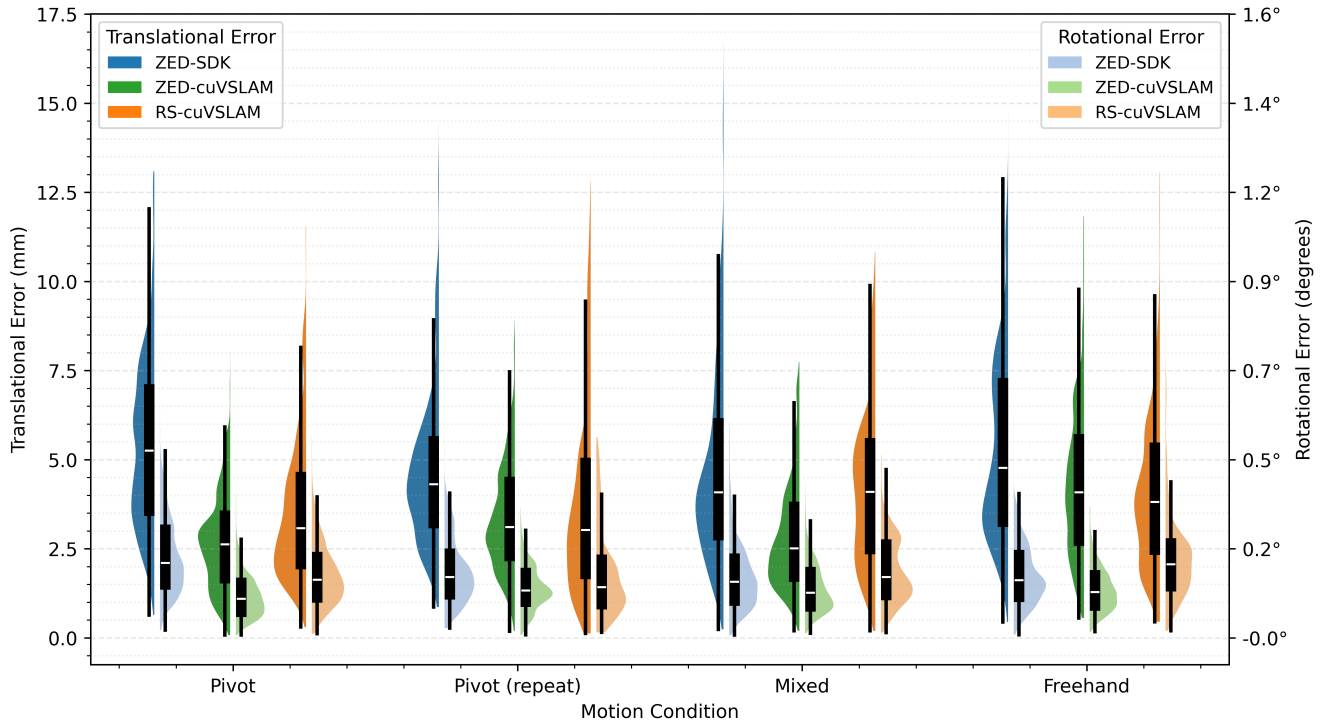


Figure 3. Hand-eye calibration error distributions across all tracking-pipeline and probe-motion type combinations. Split violin plots show translational error distributions (solid colors, left y-axis) and rotational error distributions (light colors, right y-axis).

A difference is treated as established only where the 95% bootstrap interval for the difference between the two cells excludes zero (section 2.6). Each pair of pipelines is compared on four statistics in each of three conditions: twelve comparisons per pair, thirty-six in all, of which twenty are established. ZED-cuVSLAM is ahead of ZED-SDK in ten of their twelve and ahead of RS-cuVSLAM in seven, and behind in none; the remaining three place RS-cuVSLAM ahead of ZED-SDK, all in the Pivot condition, and elsewhere that pair is unseparated and changes order between components. The established comparisons thin out as the motion is released, from ten in the Pivot condition to six in Mixed and four in Freehand, where no translational median separates at all. Two of the unresolved gaps fall below the reference floor as well as below their own interval: 0.01 mm between RS-cuVSLAM and ZED-SDK in Mixed and 0.27 mm between RS-cuVSLAM and ZED-cuVSLAM in Freehand, against the 0.28 mm contributed by the Vicon marker-cluster fit.

The repeated pivot recording provides a direct measure of trial-to-trial variability under a nominally identical protocol. Between the two pivot trials, the median rotational error changed by 0.037° for ZED-SDK (0.156° to 0.193°), 0.022° for ZED-cuVSLAM (0.121° to 0.100°) and 0.019° for RS-cuVSLAM (0.130° to 0.149°). The corresponding median translational errors changed by 0.94 mm, 0.48 mm and 0.05 mm. For most metrics this repeat variability was smaller than the variation observed between motion conditions, but not by a large margin. The rotational repeat differences correspond to 75%, 123% and 48% of the respective between-condition ranges. Differences between motion conditions that are comparable to these values, in particular for the rotational error of ZED-cuVSLAM, should therefore be treated as indicative rather than established. Across the same pair of trials the fitted camera-to-marker transformation reproduced to 1.4–2.7 mm and 0.10 – 0.12° , between a quarter and a half of its inter-condition variability, which indicates that the variability arises mainly in the estimated trajectories rather than in the calibration procedure.

The pipelines also differed in the continuity of their pose streams, in ways that were specific to each implementation. Frames for which no pose was returned were identified as gaps in the reported timestamps, and frames for which the previously reported pose was repeated unchanged were identified as stagnant. ZED-cuVSLAM returned no repeated poses in any recording, but omitted 192 of 11939 expected poses (1.61%), distributed across all four recordings. ZED-SDK and RS-cuVSLAM showed the

Table 4. Hand-eye calibration residuals for each recording and pipeline, within the shared evaluation window. Each median and 95th percentile is given with the half-width of its 95% bootstrap confidence interval (section 2.6). Bold marks the lowest value of the three pipelines in that recording and column; underlining adds that the 95% interval for its difference from each of the other two excludes zero. The final block is a second acquisition of the pivot condition, used to assess repeatability, and is not an additional motion condition. For reference, the Vicon marker-cluster fit contributes approximately 0.28 mm of translational and 0.07° of orientation uncertainty (Table 3 and text).

Recording	Pipeline	Frames	Translational [mm]		Rotational [deg]	
			Median	95th pct.	Median	95th pct.
Pivot	ZED-SDK	3304	5.26 ± 0.85	9.13 ± 1.77	0.193 ± 0.037	0.390 ± 0.044
	ZED-cuVSLAM	3253	2.63 ± 0.49	5.13 ± 0.44	0.100 ± 0.013	0.213 ± 0.028
	RS-cuVSLAM	1652	3.08 ± 0.68	7.78 ± 1.05	0.149 ± 0.021	0.307 ± 0.057
Mixed	ZED-SDK	3583	4.08 ± 0.71	9.77 ± 2.03	0.144 ± 0.031	0.318 ± 0.081
	ZED-cuVSLAM	3520	2.51 ± 0.55	5.74 ± 1.31	0.115 ± 0.016	0.237 ± 0.023
	RS-cuVSLAM	1792	4.10 ± 1.19	8.44 ± 2.06	0.156 ± 0.039	0.309 ± 0.023
Freehand	ZED-SDK	2812	4.77 ± 1.28	9.55 ± 1.77	0.148 ± 0.023	0.304 ± 0.049
	ZED-cuVSLAM	2757	4.09 ± 0.75	7.38 ± 1.57	0.117 ± 0.019	0.223 ± 0.034
	RS-cuVSLAM	1406	3.81 ± 0.63	7.66 ± 1.27	0.190 ± 0.040	0.402 ± 0.188
Pivot (repeat)	ZED-SDK	2243	4.32 ± 0.55	9.01 ± 2.37	0.156 ± 0.028	0.326 ± 0.039
	ZED-cuVSLAM	2217	3.11 ± 0.56	5.95 ± 0.71	0.121 ± 0.017	0.224 ± 0.029
	RS-cuVSLAM	1122	3.03 ± 1.23	9.39 ± 3.07	0.130 ± 0.044	0.412 ± 0.147

opposite behavior, returning a pose for essentially every frame (2 and 0 omissions, respectively) but repeating the previous pose at 54 and 40 frames within the evaluation windows. For ZED-SDK these repetitions occurred in every recording (8–21 frames each), whereas for RS-cuVSLAM they were confined to the two pivot recordings (18 and 22 frames) and absent from the mixed and freehand recordings. These two behaviors are not directly comparable, however, because the pipelines were driven differently. The ZED-SDK pipeline was processed by a script that requests each recorded frame in turn and therefore cannot omit one, whereas the cuVSLAM pipelines consumed a published stream and could drop frames under processing load. The reported omission rate is the residual after playback had already been slowed to 40% of the original rate for this reason (section 2.5), and would be expected to increase at real-time rates. Repeated poses, by contrast, are returned by the localization module itself and are a property of the pipeline rather than of how it was driven. Since the reference is interpolated to whichever timestamps a pipeline reports, omitted frames reduce the number of evaluated samples but do not otherwise bias the residuals, whereas repeated poses contribute a residual that grows with the motion occurring during the repetition.

4 DISCUSSION

As noted in section 2.6, the calibration residuals are not an error against an independent ground truth but the disagreement remaining after the SLAM trajectory is aligned to the marker motion. Two sources of uncertainty therefore limit what they can resolve: the reference itself, and the alignment to it. The reference contributes the noise of the marker-cluster fit, approximately 0.28 mm (Table 3), which bounds its translational uncertainty from below, since that residual sees only the scatter of the fit. It lies an order of magnitude below the 2.51–5.26 mm median translational errors of the pipelines (Table 4), which is what makes those residuals a practical indicator of trajectory accuracy rather than a measurement of the reference (Stolkin et al., 2006; Marchisotti and Zappa, 2021).

The same fit bounds the reference’s orientation uncertainty, but far less comfortably. That bound is not measured directly but follows from the marker noise and the geometry of the cluster. A rigid-body fit absorbs six degrees of freedom, so with five markers the 0.28 mm residual understates the noise it is fitted to, and corresponds to a per-marker position error of approximately 0.36 mm. Propagating that through a cluster of 166 mm root-mean-square radius gives an orientation uncertainty of approximately 0.07° , essentially isotropic because the cluster is close to spherically symmetric. Note that the relevant lever arm here is the extent of the marker cluster and not the camera-to-marker distance, which governs the opposite propagation. In the latter, an error of 0.07° contributes approximately 0.14 mm of position error at the camera, small compared with the marker residual itself.

The consequence is that the measured rotational errors, $0.100\text{--}0.193^\circ$, lie only a factor of 1.4 to 2.8 above this 0.07° floor, whereas the translational errors lie 9 to 19 times above their 0.28 mm one. Treating pipeline error and reference noise as independent and additive in quadrature, the 0.100° median of ZED-cuVSLAM corresponds to roughly 0.07° of pipeline error once the reference contribution is removed, while the 0.193° of ZED-SDK is barely affected. Reference noise therefore inflates the best-performing pipeline proportionally more than the worst, compressing the apparent differences between them.

The second source is the alignment itself. The marker-cluster fit noise, any error remaining in the temporal offset, and the interpolation of the reference onto the camera timestamps all enter through the estimated camera-to-marker transformation ${}^M\mathbf{T}_C$, which varies by 5.0–10.0 mm between recordings of the same rigid assembly. However, because ${}^M\mathbf{T}_C$ is fitted jointly with the world transform and the two are strongly coupled, most of a displaced lever arm is reabsorbed by a displaced world origin. Consequently, the measured effect of that spread on a median is a few tenths of a millimeter ([section 3](#)).

What a comparison can resolve then depends on whether it is made within a single recording or across recordings. Within a recording, all three pipelines are evaluated against the same reference trajectory, so the reference orientation noise enters their residuals identically and largely cancels when those residuals are differenced. The comparison is then limited by the precision of the statistic itself. Across recordings the reference and the hand-eye calibration are estimated independently for each, so neither cancels and the calibration term adds to that. Because the residual series are strongly autocorrelated, the precision of each statistic follows from effective sample sizes of 8–92 rather than from the 1400–3600 frames evaluated ([Appendix A](#)), and it varies by a factor of up to eight between the best- and worst-determined cells; each comparison is therefore judged against its own intervals rather than against a single resolution figure quoted for the study as a whole. Stating such a floor explicitly is uncommon in the probe-tracking literature, where the reference is more often characterized by a manufacturer specification or by an unquantified assertion of sub-millimeter accuracy ([Sun et al., 2014](#); [Wang et al., 2017](#); [Benjamin et al., 2020](#); [He et al., 2024](#)); without it, it is unclear which of the reported differences a given measurement could actually resolve.

Differences between the motion conditions lie almost entirely below the resolution of this comparison. Across the three conditions the median rotational error varied by 0.049° for ZED-SDK, 0.040° for RS-cuVSLAM and 0.018° for ZED-cuVSLAM, and only ZED-SDK's exceeds its own limit, by two percent. The corresponding translational ranges are 1.17, 1.02 and 1.58 mm, and ZED-SDK's again sits on its limit, of 1.14 mm. Both of ZED-SDK's margins are too small to survive the calibration term that an across-recording comparison additionally carries. ZED-cuVSLAM's translational range is the exception. Its median in the freehand recording exceeds that in the pivot recording by 1.46 mm and that in the mixed recording by 1.58 mm, against limits of 0.90 and 0.91 mm, and adding the largest calibration shift measured in [section 3](#), 1.05 mm, in quadrature raises those limits only to 1.38 and 1.39 mm. The two pivot trials differ by 0.48 mm against a limit of 0.94 mm, so trial-to-trial variation does not account for it either. With one recording per condition, though, the motion is confounded with everything else that differed between those recordings, so what this resolves is a difference between recordings and not its cause. Apart from ZED-cuVSLAM's translational range, the apparent trends do not share a common direction. Three of the six pipeline-component combinations are largest under the unconstrained freehand motion and three are not, with both components of ZED-SDK's error largest under the most constrained motion instead. For those the single repeated pivot recording shows trial-to-trial variation of comparable magnitude, and establishing motion dependence, in that case or in any other, would require repeated acquisitions of each condition.

Differences between pipelines were larger and partly established. Seventeen of the twenty established comparisons place ZED-cuVSLAM ahead and none places it behind ([section 3](#)), so its advantage is not confined to one recording or one statistic. However, these twenty comparisons are not spread evenly across the conditions and the pipeline pairs. Among the freehand translational errors only one is established and none of the medians, and ZED-SDK and RS-cuVSLAM separate only in the pivot condition. The tails establish slightly more comparisons than the medians, eleven against nine. Although a tail is estimated from fewer effective observations than a center, the differences between pipelines grow faster toward the tail than the uncertainty does. ZED-cuVSLAM's lead is therefore supported at both the center and the tail of the error distributions. The tail is also the more relevant comparison for freehand reconstruction, where the largest excursions rather than the typical error determine the worst-case misplacement of an image within the volume.

The two localization frameworks do not return the same kind of trajectory. At the end of playback cuVSLAM exports the trajectory it has been optimizing throughout, in which every pose has been refined using observations made later in the sequence. The ZED SDK's Python API exposes poses only as they are produced and offers no way to retrieve a globally optimized trajectory once playback has finished; its spatial-memory optimization improves subsequent poses but cannot revise those already reported. Part of the accuracy advantage measured for the cuVSLAM pipelines is therefore attributable to this difference rather than to frame-to-frame localization quality, and the present comparison cannot separate the two contributions. For

freehand ultrasound the distinction is nonetheless the operationally relevant one, because reconstruction is performed after the sweep is complete and a batch-optimized trajectory is therefore usable. A framework that cannot supply one forgoes an accuracy improvement that the application would otherwise accept for free. The limitation belongs to the ZED SDK as deployable through its public interface, and would not be removed by a different experimental arrangement.

Beyond tracking accuracy, the pipelines differed in the continuity of their pose streams. ZED-cuVSLAM omitted poses and never repeated one, while ZED-SDK and RS-cuVSLAM did the reverse (section 3). Both behaviors are harmful, but not equally so. An omitted pose reduces the effective sampling rate, but the gap it leaves is visible in the reported timestamps, whereas a repeated pose misreports the probe as stationary and leaves no such gap. For freehand ultrasound, where each pose places an image within a volume, that makes the repeated pose the more damaging of the two, and it is also the behavior that belongs to the localization module rather than to the offline arrangement used here. The pipelines also differed in nominal rate, RS-cuVSLAM producing poses at 30 Hz against 60 Hz for the ZED-based pipelines. Because the Vicon trajectory was acquired at 240 Hz and interpolated to each pipeline's synchronized timestamps, that difference is unlikely to have influenced the accuracy trends reported above. It bears instead on how densely a sweep is sampled, which is the property these continuity failures also degrade. Separately, the ZED-SDK pose stream proved the least reliably alignable to an external reference. Refining its temporal offset against the rotational residual moved the estimate by 5.7 ms on average and by as much as 11.8 ms, against 1.3 and 1.9 ms for the two cuVSLAM pipelines. For comparison, cross-device synchronization in instrumented freehand ultrasound has been validated at 43.0 ± 16.3 ms between the force and motion streams (Schimmoeller et al., 2020), a figure limited by the 33 Hz motion sampling rate of that arrangement.

Placing the accuracy figures beside published results requires care, because the reported metrics differ in kind. Several studies of camera-based probe tracking quote error accumulated over a stated travel distance (Sun et al., 2014; Wang et al., 2017); such rates integrate along the trajectory and are not defined over the 7–13 m paths evaluated here, so they cannot be compared numerically with the present residuals. Four results are closer in kind. A probe-mounted vision and inertial system reported 3.78 mm and 0.36° against a robot-arm reference over a single 30 s trajectory (He et al., 2024); an outside-in computer-vision approach reported a mean error norm of 2.86 mm against a ten-camera optical system (De Sanctis et al., 2024); a RealSense D435i mounted on a probe reported a trajectory RMSE of 10.40 mm against an optical tracker (Benjamin et al., 2020); and a probe-mounted camera pair localizing against a printed textured workspace reported mean errors of 1.46 mm and 1.37° over a 0.5 m path, likewise against a robot-arm reference (Zhang et al., 2025). The translational residuals reported here, 2.51–5.26 mm in the median and 5.13–9.77 mm at the 95th percentile, are of the same order as the first two and below the third, which is the closest hardware match to RS-cuVSLAM. They exceed the fourth, which was obtained over a path more than an order of magnitude shorter and by a method that recovers each pose absolutely against a known pattern rather than by integrating motion, so that error does not accumulate along the trajectory. The rotational comparison is less direct, since He et al. (2024) report the mean of three per-axis absolute errors, a quantity that understates a three-dimensional angular magnitude; the appropriate comparison is therefore with the 95th percentiles reported here, $0.213\text{--}0.402^\circ$, rather than with the medians. The rotational figure of Zhang et al. (2025) runs in the opposite direction to its translational one, exceeding those percentiles by a factor of three to six.

Locating an ultrasound frame within a volume also requires the probe-to-image-plane calibration, and the two terms contribute differently. Reported calibration accuracies are of the order of a millimeter, from 0.38 mm for a robot-assisted automated procedure (Paralika et al., 2025) and 0.66 mm for a real-time system with accuracy feedback (Chen et al., 2009) to 1.0–1.4 mm for wire-phantom methods, degrading to about 2 mm at greater imaging depth (Carbajal et al., 2013; Shen et al., 2019). The translational budget for placing a frame is therefore dominated by probe localization at the residuals reported here. The angular comparison is less favorable, because an image-plane orientation error of $0.20\text{--}0.80^\circ$ has been reported for automated calibration (Paralika et al., 2025), comparable to or larger than the rotational residuals measured here. Improving rotational tracking alone would therefore not reduce the error in frame placement.

The magnitude of these residuals can also be placed in context against published reconstruction results. Ultrasound-derived surfaces of a residual limb, reconstructed with camera-based motion compensation, have been shown to agree with magnetic-resonance-derived surfaces to within approximately 2–3 mm (Ranger et al., 2019), indicating that tracking errors of a few millimeters are compatible with usable limb geometry. The comparison is indicative rather than equivalent, however, since the motion involved was the involuntary sway of a submerged limb over acquisitions of roughly fifteen seconds and the reported agreement combines imaging, segmentation and tracking. Quantifying the relationship directly is the natural next step, and the present arrangement supports it. Acquiring ultrasound while the probe is tracked simultaneously by the reference system and by each pipeline would allow a volume to be reconstructed from each pose stream and the resulting geometries compared. Such an experiment would also supply one of the error terms that inverse identification from ultrasound indentation has been argued

to require (Fougeron et al., 2024).

4.1 Limitations

The study comprises four recordings, made by a single operator in a single session: one for each of the three motion conditions, with the pivot condition acquired twice. Because no condition other than the pivot was repeated, a difference between two conditions cannot be separated from a difference between the two recordings that carry it, and the condition-wise results are therefore reported as observations rather than as tested effects. The pivot pair provides the only estimate of trial-to-trial variability, and that variability is of the same order as the differences observed between conditions. The two pivot recordings also differ from each other in angular extent and in calibration margin (Table 2), so they repeat the protocol rather than the motion, and the variability they capture includes that difference.

The design isolates the localization algorithm in one comparison and confounds several factors in the other. ZED-SDK and ZED-cuVSLAM were run on the same recordings from the same camera, so differences between them are attributable to the localization framework alone. ZED-cuVSLAM and RS-cuVSLAM share the algorithm but differ in camera model, image resolution and modality (1920×1200 stereo against 1280×800 8-bit infrared), image rate (60 against 30 Hz), lens and stereo baseline, and IMU; differences between them cannot be attributed to any one of these factors. The parameters governing localization were left largely at their defaults, so the comparison does not represent the best accuracy each pipeline might reach after tuning. The configurations were not wholly default, however, since the cuVSLAM pipelines required adjustments to how the recorded streams were delivered before they would process them without dropping frames (section 2.5).

The pipelines evaluated here localize against the surrounding room, so a rigid displacement of the patient during a sweep would leave the estimated probe pose correct in the room frame and wrong relative to the anatomy. Approaches that reference the patient directly, by tracking skin features or external anatomy with a probe-mounted camera, reject such motion by construction (Sun et al., 2014; Wang et al., 2017). The recordings analyzed here were acquired with the cameras rigidly mounted to a frame and no patient present, so this failure mode could not arise and its magnitude is not quantified.

As set out above, no arrangement available through the ZED SDK Python API would make a causal and a batch-optimized trajectory comparable, so the comparison does not separate localization quality from trajectory optimization. The reported differences therefore characterize the pipelines as they can be deployed, but they do not isolate the accuracy of the underlying visual-inertial estimators, and they should not be extrapolated to a real-time setting in which no pipeline could apply full-sequence optimization.

Three limits apply on the analysis side, quantified earlier in this section and recalled here. The rotational comparison is close to the resolution of its own reference, which compresses the apparent differences between pipelines and does so most for the best-performing one. The hand-eye calibration was re-estimated for every recording, so each residual conflates tracking error with the error of that recording's calibration, and the two cannot be separated within a recording; the reproducibility of the fitted camera-to-marker transform bounds the second contribution but does not remove it. Finally, the last stage of temporal synchronization minimizes a rotational-alignment objective, and the rotational residual is then reported from the same series. With an effective sample size of order tens of samples per recording, the resulting optimistic bias on a one-parameter fit is bounded at a few percent, but it is not zero, and it applies unequally across pipelines because the required refinement was larger for ZED-SDK.

None of the pipelines was evaluated in real time. All processing was offline from recorded data, and the cuVSLAM pipelines required playback at 40% of the acquisition rate to limit frame loss. The 1.61% omission rate reported for ZED-cuVSLAM is therefore a lower bound on what would occur at acquisition rate on the same hardware, and the pose-continuity comparison should not be read as a statement about real-time behavior.

Finally, the study measures probe pose accuracy and not reconstruction accuracy. No ultrasound images were acquired, and the relationship between the residuals reported here and geometric error in a compounded volume was not established; that relationship depends on the sweep geometry, the image resolution, and the compounding method, none of which were varied. The clinical and biomechanical motivations set out in the introduction are therefore stated rather than closed, since this work quantifies one input to those applications and evaluates none of them. The recordings were made in a laboratory with a static high-contrast scene deliberately kept in view (Figure 1b), which is favorable for visual odometry; clinical settings with specular surfaces, moving personnel, or a probe oriented toward a patient rather than a textured wall would be expected to be more demanding. Recordings were 37 to 60 s long, so drift accumulated over the longer sweeps used in some freehand protocols was not assessed. The frame used here carries the cameras in a fixed geometry but is not an instrumented ultrasound probe, and the motion conditions were operator-defined rather than derived from clinical scanning protocols.

5 CONCLUSION

This study compared three visual–inertial tracking pipelines against a Vicon reference under three representative motion conditions. Tracking accuracy depended primarily on the selected pipeline. Variation with motion condition did not follow a common direction across pipelines, and only one pipeline showed a between-condition difference exceeding the resolution of the comparison, while a repeated acquisition of one condition showed trial-to-trial variation of similar magnitude. Establishing motion dependence at this scale would require repeated acquisitions of each condition, but the present results indicate that the choice of pipeline is the larger determinant of performance.

Among the evaluated pipelines, ZED-cuVSLAM demonstrated the strongest overall performance. It achieved the lowest median rotational error and the narrowest error distribution in all three motion conditions, and the lowest 95th-percentile error of any pipeline in every condition and both error components, which is the comparison least affected by reference uncertainty. It also produced pose estimates at 60 Hz rather than 30 Hz. For offline volumetric reconstruction, in which each ultrasound image is assigned a pose interpolated from the tracking stream, a denser pose stream is preferable, although the effect of pose sampling rate on reconstruction accuracy was not evaluated in this work. Its pose stream also contained no repeated poses, although it omitted a small fraction of frames under the offline playback arrangement used here. Taken together, these findings identify ZED-cuVSLAM as the most suitable tracking pipeline for the evaluated experimental setup.

CODE AVAILABILITY

The codes used to conduct the analyses described in this study are freely and openly available, and can be obtained at <https://github.com/SolavLab/freehand-us-tracking-eval>. This paper refers to release version v1.1.0, archived at <https://doi.org/10.5281/zenodo.23213919> (Oddes and Solav, 2026b). Every table and figure reported here is reproducible from that release and the data it ships.

DATA AVAILABILITY

The pose and marker trajectories from which all reported statistics are computed are distributed with the code release above. The raw sensor recordings behind them, comprising the ZED stereo and inertial captures and the RealSense infrared stereo and inertial captures for all four recordings, are archived separately under CC-BY-4.0 at <https://doi.org/10.5281/zenodo.22690904> (Oddes and Solav, 2026a). They are not required to reproduce any result in this paper.

ACKNOWLEDGEMENTS

This study was supported by the Israel Science Foundation, grant no. 1750/22.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

During the preparation of this manuscript, the authors used Claude (Anthropic, accessed through Claude Code) and ChatGPT (OpenAI) to support literature searching, drafting and restructuring of the text, implementation of the analysis code released with this paper, and numerical verification of the reported values against the archived data. All cited sources were independently checked by the authors against the published records. The authors reviewed, validated and critically assessed all AI-assisted outputs and take full responsibility for the content of the manuscript.

Appendix A: TEMPORAL SYNCHRONIZATION

The cameras and the Vicon system run on independent clocks, so the temporal offset between them was estimated separately for every pipeline and every recording. All stages operate on the angular-speed magnitude, which is invariant to the choice of coordinate frame and is therefore the only quantity directly comparable between the two systems before the spatial calibration of [section 2.6](#) has been established.

Angular-speed magnitude

For a trajectory sampled at times t_i with orientations $\mathbf{R}(t_i)$, the increment between successive samples is $\Delta\mathbf{R}_i = \mathbf{R}(t_{i+1})\mathbf{R}(t_i)^{-1}$, and the angular speed is its rotation angle divided by the sampling interval,

$$\omega_i = \frac{\theta(\Delta\mathbf{R}_i)}{t_{i+1} - t_i}, \quad \theta(\mathbf{R}) = \arccos\left(\frac{\text{tr}\mathbf{R} - 1}{2}\right), \quad (\text{A.1})$$

assigned to the midpoint $\frac{1}{2}(t_i + t_{i+1})$. Here $\theta(\mathbf{R}) \in [0, \pi]$ is the rotation angle of $\mathbf{R}$, equivalently the Euclidean norm of its rotation vector, which is the form used in the implementation.

This is the geodesic angular speed, and it is the appropriate quantity here. It is invariant to the coordinate frames in which the two trajectories are expressed, which is essential because those frames are unrelated until the spatial calibration has been estimated.

Stage 1: cross-correlation

The angular speed was computed for the SLAM and Vicon trajectories, the Vicon series interpolated linearly onto the camera sample times, and both series reduced to zero mean and unit variance. Their cross-correlation was evaluated at all integer sample lags and the peak located, after which the correlation curve was interpolated by a cubic spline and the peak refined to sub-sample precision by bounded minimization over an interval of one sample either side. The peak correlation exceeded 0.95 for every pipeline in every recording, so this stage does not by itself separate the pipelines. The difference in how tightly the offset is constrained emerges only from the width of the objective's minimum, quantified below. Figure A.1 shows the aligned series.

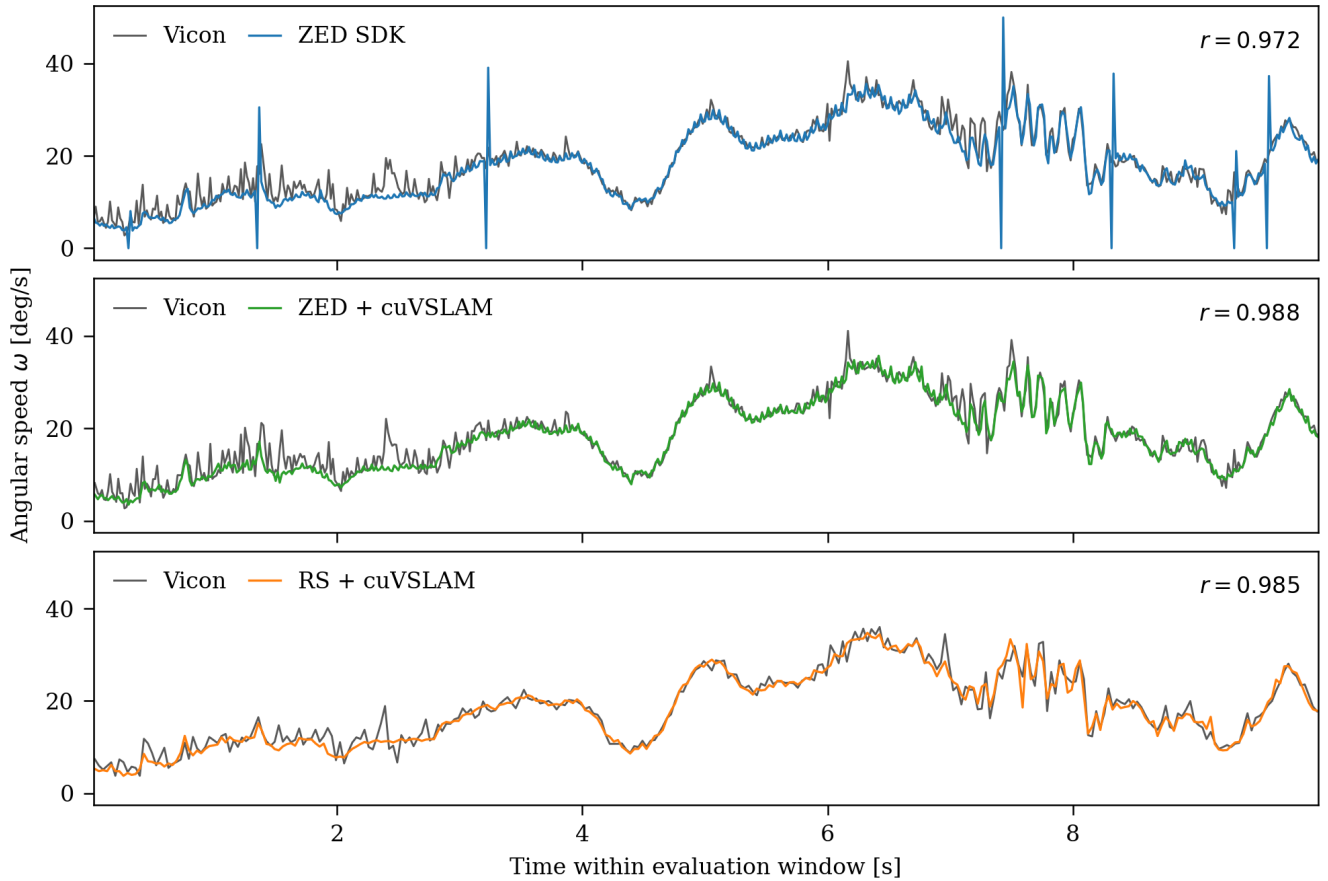


Figure A.1. Angular-speed magnitude of each pipeline against the Vicon trajectory, for a ten-second segment of the pivot recording at the estimated temporal offset. The Vicon trajectory is differentiated at its native 240 Hz rate and the resulting angular speed interpolated onto the camera timestamps, which is the order the objective uses; differentiating after resampling instead yields a visibly smoother trace, since the interval is then 17 ms rather than 4.2 ms. The correlation quoted in each panel is taken over the whole evaluation window. The excursions to zero in the ZED-SDK trace are frames at which the previous pose was repeated: the rotational increment vanishes and the following interval carries twice the rotation.

Stage 2: minimization of the angular-speed mismatch

The stage-1 estimate was refined by minimizing the root-mean-square disagreement between the two angular-speed series,

$$J(\tau) = \left(\frac{1}{|\mathcal{I}|} \sum_{t \in \mathcal{I}} [\omega_C(t) - \omega_V(t + \tau)]^2 \right)^{1/2}, \quad (\text{A.2})$$

where ω_V was resampled onto the camera midpoints by cubic spline and $\mathcal{J}$ restricts the sum to camera samples lying inside the interval spanned by the Vicon recording, so that no extrapolated value enters the objective. Minimization used the Nelder–Mead simplex (Nelder and Mead, 1965; Virtanen et al., 2020), at most 20 iterations, with a tolerance of 10^{-5} ms.

Because ω is a scalar, stages 1 and 2 constrain the offset less tightly than the pose trajectories themselves would. Measured directly on the objective, they localized the offset to approximately 3 ms for both cuVSLAM pipelines but only to approximately 12 ms for ZED-SDK.

Stage 3: refinement against the rotational residual

Once the spatial calibration was available, the offset was refined once more, by minimizing the median rotational residual $\Delta\phi$ (Equation 5) with respect to τ . For each pipeline the objective was first evaluated on a grid spanning ± 16 ms in steps of 4 ms. If the minimum fell on an edge the span was doubled, up to a limit of ± 64 ms. The interval bracketing the grid minimum was then refined by bounded scalar minimization to a tolerance of 0.25 ms, and a refined value was accepted only if it did not increase the objective.

The shared evaluation window depends on the offsets, and the refined offsets depend on the window through the interval over which the residual is evaluated. The two were therefore iterated to a fixed point, terminating when no offset moved by more than 0.5 ms or after four iterations. Convergence occurred within two iterations in every case. This stage changed the estimate by less than 3 ms for the cuVSLAM pipelines and by up to 11.8 ms for ZED-SDK. Figure A.2 shows the objective in each condition.

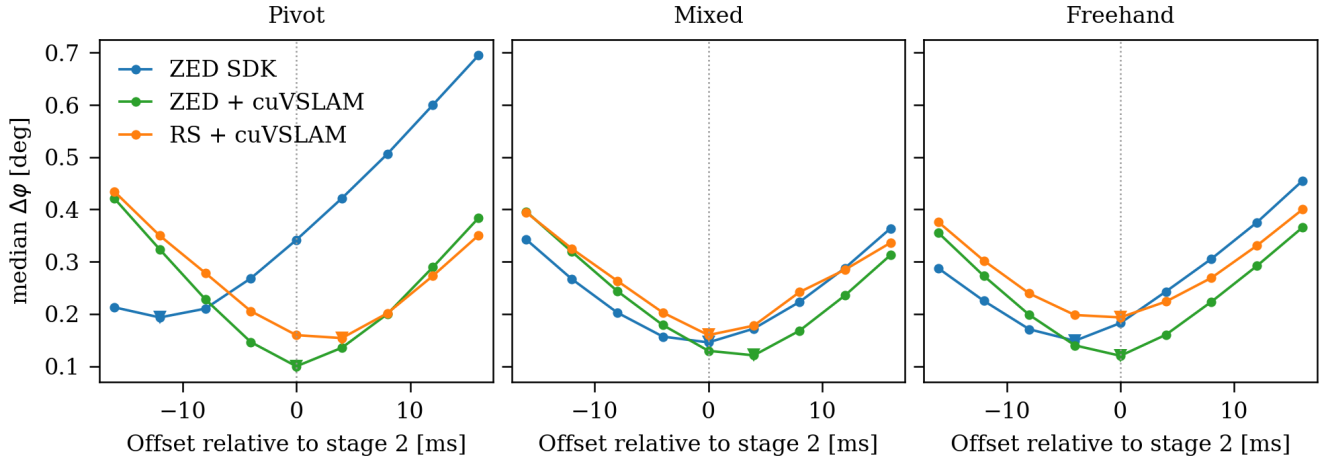


Figure A.2. Median rotational residual as a function of the temporal offset, plotted relative to the stage-2 estimate (dotted line), with triangles marking the grid minimum. For the pivot recording the ZED-SDK minimum lies 12 ms from the stage-2 estimate, whereas both cuVSLAM pipelines are within 4 ms; in the mixed and freehand recordings all three agree to within 4 ms. The ZED-SDK objective is also the shallowest on the left of its minimum in the pivot condition, which is why the scalar angular-speed criterion localizes it least tightly.

Refining a temporal offset against the same residual that is subsequently reported introduces an optimistic bias, whose scale is set by the number of independent samples rather than the number of frames. The residual series are strongly autocorrelated. Integrating the normalized autocorrelation of the per-frame rotational residual over the initial run of lags at which it remains positive gives correlation times of 0.65–2.90 s across the nine pipeline-condition cells, so the effective sample size — the recording duration divided by that time (Smith et al., 2018) — is 16–92 rather than the 1400–3600 frames evaluated, placing the expected optimism of a one-parameter fit at a few percent. This is a scaling argument rather than a bound, but it suffices here, since the improvement obtained for ZED-SDK in the pivot condition was 44%, an order of magnitude larger. The translational residuals are generally more strongly autocorrelated still, giving effective sample sizes of 8–42 over the same cells. These figures, not the frame counts, govern the precision of the statistics reported in the main text. Standard errors there are obtained by the moving-block bootstrap of section 2.6, which takes its block length from the correlation time measured here.

Interpolation of the Vicon trajectory

With the offset fixed, the Vicon marker trajectories were resampled to the camera timestamps by linear interpolation, and a rigid body was fitted to the five markers at each resampled instant by the method of section 2.2. Interpolating the marker

positions and then fitting, rather than fitting and then interpolating the resulting poses, matters because the rigid-body fit is nonlinear. Cubic interpolation is used only for the angular-speed series entering the stage-2 objective (Equation A.2), where linear interpolation of a 240 Hz trajectory introduces a ripple at the Vicon sample period.

Appendix B: WELL-POSEDNESS OF THE HAND-EYE CALIBRATION

This appendix derives the margin $1 - \sigma_2$ reported in Table 2 and used in section 2.6 to establish that the calibration was well determined in every recording.

The operator that Shah's method factorizes

Write $\mathbf{R}_A(t)$, $\mathbf{R}_B(t)$, $\mathbf{R}_X$ and $\mathbf{R}_Z$ for the rotational parts of ${}^C\mathbf{T}_W(t)$, ${}^M\mathbf{T}_V(t)$, ${}^C\mathbf{T}_M$ and ${}^W\mathbf{T}_V$. The calibration states that

$$\mathbf{R}_A(t) \mathbf{R}_Z = \mathbf{R}_X \mathbf{R}_B(t) \quad (\text{B.1})$$

at every instant. Right-multiplying by $\mathbf{R}_B(t)^\top$, which is its inverse, gives $\mathbf{R}_X = \mathbf{R}_A(t) \mathbf{R}_Z \mathbf{R}_B(t)^\top$. Applying the identity $\text{vec}(\mathbf{AMB}) = (\mathbf{B}^\top \otimes \mathbf{A}) \text{vec}(\mathbf{M})$, with column-major vec , yields

$$\text{vec}(\mathbf{R}_X) = (\mathbf{R}_B(t) \otimes \mathbf{R}_A(t)) \text{vec}(\mathbf{R}_Z). \quad (\text{B.2})$$

Every instant therefore contributes a 9×9 matrix that carries the same vector $\text{vec}(\mathbf{R}_Z)$ to the same image $\text{vec}(\mathbf{R}_X)$, and so does their average,

$$\mathbf{K} = \frac{1}{N} \sum_{i=1}^N \mathbf{R}_B(t_i) \otimes \mathbf{R}_A(t_i). \quad (\text{B.3})$$

This is, up to the factor N^{-1} , the matrix whose singular value decomposition Shah's method computes (Shah, 2013). The reference implementations omit that factor, which rescales the singular values without affecting the singular vectors from which the calibration is recovered. It is retained here because the singular values themselves are the quantity of interest. Note also that Shah (2013) writes the problem as $\mathbf{AX} = \mathbf{YB}$, so that the transformation denoted $\mathbf{X}$ there corresponds to $\mathbf{R}_Z$ here and $\mathbf{Y}$ to $\mathbf{R}_X$.

The largest singular value is unity

Each term of Equation B.3 is a Kronecker product of two rotations and is therefore orthogonal, so each preserves length. An average of length-preserving maps is contractive except along directions on which every term agrees, since it is only there that no cancellation occurs. Because $\mathbf{R}_X$ and $\mathbf{R}_Z$ are constant, Equation B.2 shows that $\text{vec}(\mathbf{R}_Z)$ is such a direction, since every term sends it to the same image. Both vectors have length $\sqrt{3}$, since $\|\mathbf{R}\|_F^2 = \text{tr}(\mathbf{R}^\top \mathbf{R}) = 3$ for any rotation. Hence

$$\sigma_1(\mathbf{K}) = 1, \quad (\text{B.4})$$

attained with $\text{vec}(\mathbf{R}_Z)/\sqrt{3}$ and $\text{vec}(\mathbf{R}_X)/\sqrt{3}$ as the corresponding right and left singular vectors, so that the leading singular pair is the calibration itself.

Degeneracy and the margin

A second unit singular value would mean a second direction on which every term of Equation B.3 agrees, and therefore a second pair of transformations consistent with the data. The calibration would not be identifiable. The gap

$$m = 1 - \sigma_2(\mathbf{K}) \quad (\text{B.5})$$

therefore measures how sharply the true solution is separated from the nearest competing one, with $m = 0$ marking a degenerate motion. The condition is the algebraic counterpart of the familiar requirement that the motion rotate about more than one axis. If every $\mathbf{R}_B(t_i)$ shares a common axis $\mathbf{n}$, then $\mathbf{nn}^\top$ is invariant under conjugation by all of them and supplies exactly such a second direction.

The margin depends only on the reference trajectory

Substituting $\mathbf{R}_A = \mathbf{R}_X \mathbf{R}_B \mathbf{R}_Z^\top$, which is Equation B.1 rearranged, and using $(\mathbf{AC}) \otimes (\mathbf{BD}) = (\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D})$,

$$\mathbf{K} = (\mathbf{I} \otimes \mathbf{R}_X) \left[\frac{1}{N} \sum_{i=1}^N \mathbf{R}_B(t_i) \otimes \mathbf{R}_B(t_i) \right] (\mathbf{I} \otimes \mathbf{R}_Z^\top). \quad (\text{B.6})$$

The outer factors are orthogonal and leave the singular values unchanged, so m is a property of the Vicon trajectory alone and is consequently identical for all three pipelines within a recording.

REFERENCES

- Adriaans, C. A., Wijkhuizen, M., van Karnenbeek, L. M., Geldof, F., and Dashtbozorg, B. (2024). Trackerless 3D Freehand Ultrasound Reconstruction: A Review. *Applied Sciences*, 14(17):7991. Number: 17.
- Alapetite, A., Wang, Z., Hansen, J. P., Zajączkowski, M., and Patalan, M. (2020). Comparison of Three Off-the-Shelf Visual Odometry Systems. *Robotics*, 9(3):56. Number: 3.
- Benjamin, A., Chen, M., Li, Q., Chen, L., Dong, Y., Carrascal, C. A., Xie, H., Samir, A. E., and Anthony, B. W. (2020). Renal Volume Estimation Using Freehand Ultrasound Scans: An *Ex Vivo* Demonstration. *Ultrasound in Medicine & Biology*, 46(7):1769–1782.
- Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Cao, K., Hu, Q., Zhao, P., Ai, M., Huang, S., Li, J., and Li, J. (2026). A strict and high-accuracy time synchronization approach for consumer-grade cameras and laser scanning sensor using robust attitude feature matching in time series. *Measurement Science and Technology*, 37(5):055901.
- Carbajal, G., Lasso, A., Gómez, Á., and Fichtinger, G. (2013). Improving N-wire phantom-based freehand ultrasound calibration. *International Journal of Computer Assisted Radiology and Surgery*, 8(6):1063–1072.
- Chen, T. K., Thurston, A. D., Ellis, R. E., and Abolmaesumi, P. (2009). A Real-Time Freehand Ultrasound Calibration System with Automatic Accuracy Feedback and Control. *Ultrasound in Medicine & Biology*, 35(1):79–93.
- De Sanctis, L., Carnevale, A., Antonacci, C., Faiella, E., Schena, E., and Longo, U. G. (2024). Six-Degree-of-Freedom Freehand 3D Ultrasound: A Low-Cost Computer Vision-Based Approach for Orthopedic Applications. *Diagnostics*, 14(14):1501. Number: 14.
- Doherty, S., Landis, B., Owings, T. M., and Erdemir, A. (2022). Template models for simulation of surface manipulation of musculoskeletal extremities. *PLoS ONE*, 17(8):e0272051.
- Fattori, G., Lomax, A. J., Weber, D. C., and Safai, S. (2021). Technical assessment of the NDI Polaris Vega optical tracking system. *Radiation Oncology*, 16(1):87.
- Fougeron, N., Oddes, Z., Ashkenazi, A., and Solay, D. (2024). Identification of constitutive materials of bi-layer soft tissues from multimodal indentations. *Journal of the Mechanical Behavior of Biomedical Materials*, 155:106572.
- Fougeron, N., Rohan, P.-Y., Haering, D., Rose, J.-L., Bonnet, X., and Pillet, H. (2020). Combining Freehand Ultrasound-Based Indentation and Inverse Finite Element Modeling for the Identification of Hyperelastic Material Properties of Thigh Soft Tissues. *Journal of Biomechanical Engineering*, 142(9):091004.
- Gebhardt, C., Göttling, L., Buchberger, L., Ziegler, C., Endres, F., Wuermeling, Q., Holzapfel, B. M., Wein, W., Wagner, F., and Zettinig, O. (2023). Femur reconstruction in 3D ultrasound for orthopedic surgery planning. *International Journal of Computer Assisted Radiology and Surgery*, 18(6):1001–1008.
- Habersack, A., Zussner, T., Thaller, S., Tilp, M., Svehlik, M., and Kruse, A. (2022). Validity and reliability of a novel 3D ultrasound approach to assess static lengths and the lengthening behavior of the gastrocnemius medialis muscle and the Achilles tendon in vivo. *Knee Surgery, Sports Traumatology, Arthroscopy*, 30(12):4203–4213.
- He, W., Zhao, B., Zhou, Y., Wu, R., Wu, G., Li, Y., Lu, M., Zhu, L., and Gao, Y. (2024). Freehand 3D Ultrasound Imaging Based on Probe-mounted Vision and IMU System. *Ultrasound in Medicine & Biology*, 50(8):1143–1154.
- Koivukangas, T., Katisko, J. P., and Koivukangas, J. P. (2013). Technical accuracy of optical and the electromagnetic tracking systems. *SpringerPlus*, 2(1).
- Korovko, A., Slepichev, D., Efitov, A., Dzhumamuratova, A., Kuznetsov, V., Rabeti, H., Biswas, J., and Pouya, S. (2025). cuvsam: Cuda accelerated visual odometry and mapping. *arXiv preprint arXiv:2506.04359*.
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241.
- Lasso, A., Heffter, T., Rankin, A., Pinter, C., Ungi, T., and Fichtinger, G. (2014). Plus: Open-source toolkit for ultrasound-guided intervention systems. *IEEE Transactions on Biomedical Engineering*, 61(10):2527–2537.
- Lee, M. W. (2014). Fusion imaging of real-time ultrasonography with CT or MRI for hepatic intervention. *Ultrasonography*, 33(4):227–239.
- Marchisotti, D. and Zappa, E. (2021). Virtual simulation benchmark for the evaluation of simultaneous localization and mapping and 3D reconstruction algorithm uncertainty. *Measurement Science and Technology*, 32(9):095404.
- Maybody, M., Stevenson, C., and Solomon, S. B. (2013). Overview of Navigation Systems in Image-Guided Interventions. *Techniques in Vascular and Interventional Radiology*, 16(3):136–143.
- Mercier, L., Langø, T., Lindseth, F., and Collins, D. L. (2005). A review of calibration techniques for freehand 3-d ultrasound systems. *Ultrasound in Medicine & Biology*, 31(4):449–471.
- Moerman, K. M., Solay, D., Sengeh, D. M., and Herr, H. M. (2019). Automated and Data-driven Computational Design of Patient-Specific Biomechanical Interfaces. *engrXiv preprint*, pages 1–16.
- Mozaffari, M. H. and Lee, W.-S. (2017). Freehand 3-D Ultrasound Imaging: A Systematic Review. *Ultrasound in Medicine &*

- Biology*, 43(10):2099–2124.
- Nelder, J. A. and Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313.
- Oddes, Z. and Solav, D. (2023). Identifiability of soft tissue constitutive parameters from in-vivo macro-indentation. *Journal of the Mechanical Behavior of Biomedical Materials*, 140:105708.
- Oddes, Z. and Solav, D. (2026a). Freehand 3D ultrasound probe tracking — raw sensor recordings (ZED SVO2 + RealSense bag). Dataset, CC-BY-4.0. DOI: 10.5281/zenodo.22690904.
- Oddes, Z. and Solav, D. (2026b). vipose: visual-inertial pose tracking evaluation. Software, version 1.1.0, MIT licence. DOI: 10.5281/zenodo.23213919.
- Paralikar, A., Mantripragada, P., Nguyen, T., Arjoune, Y., Shekhar, R., and Monfaredi, R. (2025). Robot-assisted ultrasound probe calibration for image-guided interventions. *International Journal of Computer Assisted Radiology and Surgery*, 20(5):859–868.
- Peng, C., Cai, Q., Chen, M., and Jiang, X. (2022). Recent Advances in Tracking Devices for Biomedical Ultrasound Imaging Applications. *Micromachines*, 13(11):1855. Number: 11.
- Ranger, B. J., Feigin, M., Zhang, X., Moerman, K. M., Herr, H., and Anthony, B. W. (2019). 3D Ultrasound Imaging of Residual Limbs With Camera-Based Motion Compensation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(2):207–217. Conference Name: IEEE Transactions on Neural Systems and Rehabilitation Engineering.
- Ranger, B. J., Moerman, K. M., Anthony, B. W., and Herr, H. M. (2023). Constitutive parameter identification of transtibial residual limb soft tissue using ultrasound indentation and shear wave elastography. *Journal of the Mechanical Behavior of Biomedical Materials*, 137:105541.
- Sabatini, A. M. (2006). A wavelet-based bootstrap method applied to inertial sensor stochastic error modelling using the Allan variance. *Measurement Science and Technology*, 17(11):2980–2988.
- Sankepalle, D. M., Anthony, B., and Mallidi, S. (2023). Visual inertial odometry enabled 3D ultrasound and photoacoustic imaging. *Biomedical Optics Express*, 14(6):2756–2772.
- Savitzky, A. and Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8):1627–1639.
- Schenker, N. and Gentleman, J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician*, 55(3):182–186.
- Schimmoller, T., Colbrunn, R., Nagle, T., Lobosky, M., Neumann, E. E., Owings, T. M., Landis, B., Jelovsek, J. E., and Erdemir, A. (2019). Instrumentation of off-the-shelf ultrasound system for measurement of probe forces during freehand imaging. *Journal of Biomechanics*, 83:117–124.
- Schimmoller, T., Neumann, E. E., Owings, T. M., Nagle, T. F., Colbrunn, R. W., Landis, B., Jelovsek, J. E., Hing, T., Ku, J. P., and Erdemir, A. (2020). Reference data on in vitro anatomy and indentation response of tissue layers of musculoskeletal extremities. *Scientific Data*, 7(1):20.
- Schmedling, A., Macho, E., Valenzuela, R., Diego, P., Campa, F. J., Diez, M., Corral, J., Herrero, S., and Pinto, C. (2026). Configuration evaluation and controlled validation of a multi-camera markerless human motion capture system based on ZED 2i. *Measurement Science and Technology*, 37(32):325407.
- Sengeh, D. M., Moerman, K. M., Petron, A., and Herr, H. (2016). Multi-material 3-D viscoelastic model of a transtibial residuum from in-vivo indentation and MRI data. *Journal of the Mechanical Behavior of Biomedical Materials*, 59:379–392.
- Shah, M. (2013). Solving the Robot-World/Hand-Eye Calibration Problem Using the Kronecker Product. *Journal of Mechanisms and Robotics*, 5(031007).
- Shen, C., Lyu, L., Wang, G., and Wu, J. (2019). A method for ultrasound probe calibration based on arbitrary wire phantom. *Cogent Engineering*, 6(1):1592739. _eprint: <https://doi.org/10.1080/23311916.2019.1592739>.
- Smith, B. L., Neal, D. R., Feero, M. A., and Richards, G. (2018). Assessing the limitations of effective number of samples for finding the uncertainty of the mean of correlated data. *Measurement Science and Technology*, 29(12):125304.
- Stolkin, R., Greig, A., and Gilby, J. (2006). A calibration system for measuring 3D ground truth for validation and error analysis of robot vision algorithms. *Measurement Science and Technology*, 17(10):2721–2730.
- Sun, S.-Y., Gilbertson, M., and Anthony, B. W. (2014). Probe Localization for Freehand 3D Ultrasound by Tracking Skin Features. In Golland, P., Hata, N., Barillot, C., Hornegger, J., and Howe, R., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, pages 365–372, Cham. Springer International Publishing.
- Theunissen, R., Di Sante, A., Riethmuller, M. L., and Van den Braembussche, R. A. (2008). Confidence estimation using dependent circular block bootstrapping: application to the statistical analysis of PIV measurements. *Experiments in Fluids*, 44(4):591–596.
- Tümer, N., Kok, A. C., Vos, F. M., Streekstra, G. J., Askeland, C., Tuijthof, G. J. M., and Zadpoor, A. A. (2018). Three-Dimensional Registration of Freehand-Tracked Ultrasound to CT Images of the Talocrural Joint. *Sensors*, 18(7):2375. Number: 7.

- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272.
- Wang, J., Che, C., Galeotti, J., Horvath, S., Gorantla, V., and Stetten, G. (2017). Ultrasound Tracking Using ProbeSight: Camera Pose Estimation Relative to External Anatomy by Inverse Rendering of a Prior High-Resolution 3D Surface Map. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 825–833.
- Wang, Y., Spangler, C., Tai, B. L., and Shih, A. J. (2013). Positional accuracy and transmitter orientation of the 3D electromagnetic tracking system. *Measurement Science and Technology*, 24(10):105105.
- Weide, G., Van Der Zwaard, S., Huijing, P. A., Jaspers, R. T., and Harlaar, J. (2017). 3D Ultrasound Imaging: Fast and Cost-effective Morphometry of Musculoskeletal Tissue. *Journal of Visualized Experiments*, (129):55943.
- Zhang, Y., Huang, D., Meng, M. Q.-H., Navab, N., and Jiang, Z. (2025). Freehand 3-D Ultrasound Imaging: Sim-in-the-Loop Probe Pose Optimization via Visual Servoing. *IEEE/ASME Transactions on Mechatronics*, pages 1–13.