Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Dual-Modal Fiducial Marker with Single-View X-ray Pose Recoverability for MR-Guided Surgical Navigation

Ganzhe Xu
Xiangya Hospital of Central South University
Xi Li
Xiangya Hospital of Central South University
Wenyi Xiong
Central South University
Zhen Yin
Xiangya Hospital of Central South University
Yaxin Yang
Xiangya Hospital of Central South University
Xiaopeng Wang
Xiangya Hospital of Central South University
Da Zhong
Xiangya Hospital of Central South University
Hua Liu
Xiangya Hospital of Central South University
Yang Xie
Xiangya Hospital of Central South University
Kunli Chen
Xiangya Hospital of Central South University
Chenggong Wang
wangchenggong@csu.edu.cn
Xiangya Hospital of Central South University

Article

Keywords: Surgical navigation, fiducial marker, dual-modal registration, mixed reality, pose estimation

Posted Date: July 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10110471/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on November 13th, 2026. See the published version at https://doi.org/10.1038/s41598-026-77300-3.

Abstract

Traditional surgical navigation markers rely on a fixed external reference frame. However, in mixed reality (MR) surgical navigation, the tracking camera moves with the surgeon's head, making conventional markers unsuitable for this application. This study proposes that markers for MR navigation must satisfy three design constraints: omnidirectional recognizability, X-ray recognizability, and multimodal registration.Based on these constraints, this study designs a marker structure consisting of a central cube and four cylindrical axes with varying diameters. Through mathematical analysis, it is proven that the marker’s 6 degrees of freedom can be determined from a single X-ray image under a known X-ray projection system. A pose estimation algorithm combining geometric optimization and deep learning-based feature extraction was used to validate the marker’s X-ray and optical poses.In the validation performed on 100 sets of X-ray images and 194 sets of optical images, the X-ray recognition error was measured as a coronal plane deviation of 1.584 ± 1.256° and a sagittal plane deviation of 1.898 ± 1.502°. The optical recognition success rate reached 90.2%, with an overall angular error of 1.521 ± 0.976°. The error distribution across modalities was symmetrically zero-centered, and the redundancy of the four cylindrical axes maintained solvability even under extreme viewing angles. The results indicate that this design methodology can serve as a reference for the development of MR surgical navigation markers.

1. Introduction

1.1 Real-World Challenges

The global volume of orthopedic surgery is growing explosively. In the United States alone, the annual number of joint replacement procedures has already exceeded one million, and the growth rate of some procedures has far surpassed earlier expectations. According to the latest projections, by 2040 the annual number of primary total hip and total knee arthroplasties in the United States will reach approximately 1.43 million and 3.42 million cases, respectively.[1–4]Structural factors driving this growth include global population aging, rising obesity rates, and continuously increasing patient expectations for postoperative mobility and quality of life.[5]Against this backdrop, the requirements for the alignment accuracy of implants or bone fragments in orthopedic surgery are becoming increasingly stringent. In total hip arthroplasty, an acetabular cup angular deviation greater than 10° can significantly increase the risks of dislocation and revision, while cortical breach by a pedicle screw may result in neurovascular injury.[6, 7]

Surgical navigation systems provide surgeons with real-time, quantitative three-dimensional localization by spatially registering preoperative and/or intraoperative images to the patient’s actual anatomy. They have been shown to improve implant placement accuracy, reduce intraoperative radiation exposure, and shorten the learning curve.[8]However, the information presentation paradigm of current commercial navigation systems has important limitations. Anatomical position, angle, and depth data are typically displayed on external two-dimensional monitors, requiring surgeons to repeatedly shift their gaze between the surgical field and the screen. This “attention shift” imposes a substantial cognitive burden:

surgeons must mentally transform abstract numerical information into operative actions in three-dimensional space. This cognitive process itself constitutes a potential source of error, particularly in the presence of complex anatomy or intraoperative bleeding.

1.2 Current Status of Research in China and Abroad

Mixed reality (MR) head-mounted displays (HMDs), such as the Microsoft HoloLens 2, offer a distinctly different paradigm for surgical navigation. Surgical trajectories, resection boundaries, and critical anatomical structures can be directly superimposed onto the patient's body surface in the form of three-dimensional holographic images, thus addressing the problem of "attention shift."[8, 9]Recent systematic reviews have shown that AR/MR-HMD navigation for pedicle screw placement has achieved a clinical accuracy of 98–99% (Gertzbein–Robbins grades A + B), with performance comparable to that of robotic-assisted systems costing several times more.[9–11]MR surgical navigation is widely regarded as an important direction for the next generation of navigation technology.[12]

However, in the studies mentioned above, MR technology has mainly been used for the visualization of patient information and anatomical models.[8, 12]When researchers attempt to precisely align digital twin models with the surgical field, the registration error increases sharply. Multiple teams have reported unacceptable error levels in both skeletal models and preclinical studies, which appears to be related to the direct "transplantation" of conventional navigation workflows onto HMD platforms.[13]This challenge reveals that the core issue lies not in the display interface, but in the registration process: conventional markers were never designed with the physical constraints of HMDs in mind.

Specifically, existing surgical navigation markers can generally be divided into three categories: (1) infrared reflective spheres (e.g., the passive spherical markers used in the NDI Polaris system), which rely on external fixed trackers; the sphere base creates unavoidable optical blind spots, and the markers can only be tracked within a limited solid angle.[10, 14–16](2) Planar visual markers (e.g., ArUco or AprilTag fiducial codes): they can only be detected and recognized when viewed near the front-facing normal direction, fail completely when viewed from the side or rear, and do not produce any recognizable features in X-ray images.[17, 18](3) Polyhedral coverage schemes (e.g., attaching QR-code markers to the faces of a cube): although they improve angular coverage, they still appear as homogeneous geometric shadows under X-ray imaging and therefore cannot provide pose constraints.[11, 19, 20]The common limitation of all three types of markers is that none simultaneously satisfies the requirements of HMD-based navigation for omnidirectional visibility, X-ray recognizability, and multimodal geometric correspondence.

1.3 Core Theoretical Innovation

The root cause of the above difficulties lies in two fundamental differences in physical configuration between HMD-based navigation and conventional optical navigation. These differences directly impose an entirely new set of constraints on marker design.In conventional optical surgical navigation, a fixed reference tracker is positioned outside the surgical field, providing a stable "God's-eye-view" coordinate system, while all localization markers maintain an unobstructed line of sight to the tracker.This setup

leads to two key characteristics: (1) the line-of-sight convergence zone is spatially separated from the surgeon's working area; and (2) once the initial patient-to-image registration is completed using a registration probe or preoperative imaging, the registration relationship remains valid because the tracker does not move.Therefore, conventional markers, such as NDI optical markers, only need to be designed for unidirectional, occlusion-resistant optical tracking from a fixed viewpoint (Fig. 1a).

By contrast, the use of an MR HMD for surgical navigation radically changes this physical setup. The tracking camera is now mounted on the surgeon's head, introducing two major changes.Firstly, the line-of-sight convergence zone becomes multidirectional and highly overlapping with the operative field, as the surgeon and assistant may observe the marker from virtually any angle while moving around the patient.Secondly, the reference frame is no longer stationary: the pose of the HMD changes continuously with head motion, and the relative transformation between the HMD and the patient anatomy drifts over time. We therefore derive three design constraints that markers for MR navigation must satisfy simultaneously:

Constraint 1—Omnidirectional recognizability: the marker must be identifiable and allow pose estimation over a wide range of viewing angles—ideally across the full solid angle except for the attachment surface. This is necessary because multiple HMDs may observe the same marker simultaneously from different directions, while the line of sight may be intermittently occluded by surgical instruments or the assistant's hands.

Constraint 2—X-ray recognizability: because no stationary external reference tracker is available, the relative pose between the marker and the bone must be updateable intraoperatively. Intraoperative X-ray imaging (C-arm fluoroscopy) is the most readily available imaging modality in the operating room; therefore, the marker must provide sufficient geometric features to support accurate pose estimation from a single X-ray projection.[21, 22]

Constraint 3—Multimodal registration: the marker should incorporate multiple distinguishable geometric primitives, such as points, lines, or planes, so that they can be stably and consistently extracted in both optical and X-ray images. This enables robust cross-modal correspondence under partial occlusion or measurement noise and reduces registration ambiguity.[23, 24]

No conventional marker simultaneously satisfies all three constraints: retroreflective spheres fail to meet Constraint 1, due to the blind zone introduced by the base, and Constraint 2, because they provide no distinctive features under X-ray; planar QR-code markers fail to meet Constraint 1, because they are only reliably recognizable from the front, and also fail Constraint 2; polyhedral cube-based designs fail to meet Constraint 2 as well.These three constraints constitute the first-principles foundation for marker design in MR surgical navigation and serve as the theoretical basis for the subsequent geometric design, algorithm development, and experimental validation presented in this work. To the best of our knowledge, this is the first study to systematically derive marker design requirements from the physical constraints of HMD-based navigation.

1.4 Contributions and Assumptions of This Work

Inspired by prior studies on single-view pose estimation of cylindrical objects under perspective projection[23, 24] ,we hypothesize that a simple geometric object composed of a cube and four cylindrical axes with unequal diameters can simultaneously satisfy the three constraints of omnidirectional recognizability, X-ray recognizability, and multimodal registration.The cube provides rich optical features for visible-light tracking by the HMD. The four cylinders have different diameters, so their X-ray projections can be distinguished by width, and their known radii provide scale information that removes the ambiguity of perspective projection. Because the two parts are rigidly connected, the optical and X-ray poses are related by a fixed transformation.Based on this marker, paired simulated optical and X-ray image datasets can be constructed to train a deep learning model to regress the 6-DoF pose from a single image (Fig. 1c).The above hypothesis is intended to enable the pose-transfer chain: X-ray imaging acquires the relative pose between the marker and the connection region; optical imaging acquires the relative pose between the HMD and the marker; and the HMD thus indirectly determines its relative pose with respect to the connection region. This paper focuses on validating the core prerequisite of this chain, namely, that the marker can be stably recognized under both imaging modalities.

This paper develops four lines of work around this hypothesis: (1) designing a marker geometry that satisfies the three major constraints (Fig. 2), and providing an analytical proof and algorithmic framework showing that its 6-DoF pose can be uniquely determined from a single X-ray image;(2) constructing paired simulated optical and X-ray image datasets containing labeled samples with known ground-truth poses; (3) training machine learning models to estimate the marker's rotation and translation parameters from a single X-ray image or optical image; and (4) experimentally evaluating the marker's bimodal registration accuracy on a regular model, together with consistency validation against the imposed constraints.

2. Method

2.1 Geometric Design of the Marker and Principle of Pose Estimation

The three design constraints introduced in Section 1 directly determine the geometric form of the marker.

The central cube corresponds to Constraint 1 (omnidirectional recognizability). Its six faces provide optical features from multiple viewing directions.

The four cylinders $(\mathrm{r}_1<\mathrm{r}_2<\mathrm{r}_3<\mathrm{r}_4)$ with distinct diameters correspond to Constraint 2 (X-ray recognizability). Extending from the center of the cube toward the four vertex directions of a regular tetrahedron, the cylinders are arranged such that the angle between any pair of directions is

approximately 109°, and any three directions are linearly independent. Consequently, at least three cylinders are guaranteed to be observable in an X-ray image.

Rigid connection corresponds to Constraint 3 (multimodal registration). The cube and the four cylinders are rigidly connected to form a non-deformable whole, ensuring a definite and known spatial transformation relationship between the optical pose and the X-ray pose.

Under the above geometric configuration, estimating the 6-DoF pose of the marker from a single X-ray projection image can be formulated as an overdetermined geometric constraint-solving problem.

Each cylinder provides three independent measurements in the X-ray image: the direction of its projected centerline (two degrees of freedom, comprising position and orientation) and its projected width (one degree of freedom). Together, the four cylinders yield a total of 12 measurements, which are used to solve for 6 unknown parameters (3 for rotation and 3 for translation). This overdetermined formulation $(12>6)$ensures the existence of a unique least-squares solution even in the presence of image noise.

Pose estimation is performed in two steps. First, the rotation matrix is determined using the vanishing-point directional constraints derived from the projected centerlines of the four cylinders.Second, with the rotation fixed, the translation vector is recovered from the relationship between the projected widths and the known cylinder radii.The unequal radii play two critical roles in this process. First, they provide a criterion for resolving reflection ambiguity: the width ratios produced by a mirror-symmetric solution are inconsistent with the known ordering of the cylinder radii and can therefore be rejected after translation recovery. Second, they provide an absolute scale in the second step, eliminating the translation ambiguity associated with the coupled “object scaling + distance scaling” effect.

The regular tetrahedral configuration of the marker guarantees robustness. For any viewing direction, at most one of the four cylinders can be parallel to the optical axis, in which case its projection degenerates into an ellipse whose centerline cannot be reliably extracted. Nevertheless, the remaining three cylinders still provide sufficient directional constraints, since three non-coplanar directions are enough to uniquely determine the rotation matrix. Complete analytical derivations, proofs of uniqueness, and robustness analysis are provided in the Supplementary Information Section 1.

2.2 Design of the Machine-Learning-Based Pose Estimation Algorithm

Considering the convenience and cost-effectiveness of algorithm development, this study adopts a machine-learning-based approach for pose estimation. In this section, we establish machine-learning-based pose estimation algorithms for the marker under both X-ray and optical imaging modalitiesThe proposed algorithm adopts a cascaded framework that combines learnable visual perception with explicit geometric constraints. A deep neural network is used to extract pose-related local observations from the images, including the marker target region, the three tip keypoints, and rotation priors. Based on

the known 3D model, camera intrinsic parameters, and rigid-body transformation constraints, the geometric module then recovers the final 6-DoF pose.

The X-ray branch takes the ROI of fluoroscopic images as input and utilizes a ResNet50/HopeNet pose regression network to estimate rotational priors, which are then combined with cylindrical projection geometric constraints to recover the translation. The optical branch employs a two-stage perception process: YOLOv11 detects the bounding box of the marker, while a U-Net heatmap network predicts the keypoint at the tip. Meanwhile, a parallel pose regression module provides rotational priors. Finally, the Levenberg-Marquardt optimizer estimates the pose by minimizing the reprojection error under positive depth constraints.[25–27]

Compared with methods that directly regress pose parameters in an end-to-end manner, this study restricts the visual recognition task to geometrically meaningful entities that can be stably annotated and reliably verified, enabling quality control of the network outputs through reprojection error, positive-depth constraints, and rigid-body consistency.

The complete network architecture, loss function design, training hyperparameters, and data augmentation protocol are detailed in Supplementary Information Section 2.

2.3 Data Acquisition

X-ray data acquisition was conducted on a GE Definium 6000 system. To accurately obtain the ground-truth 6-DoF pose of the marker in three-dimensional space, this study designed an IMU-based pose reference frame fixed with Kirschner wires.

Optical data acquisition was performed to simulate marker observation from the perspective of a mixed reality (MR) head-mounted display. The experiment was conducted under controlled indoor lighting conditions, using a Canon R7 camera as the optical sensor.To establish a spatial coordinate reference, 6 cm × 6 cm printed black-and-white checkerboard sheets were laid on both the tabletop and the floor as reference.the center point of the marker was determined to be at $(1,8)$. The plumb line was also used to determine the ground-projected coordinate position of the center point of the camera tripod, thus eliminating systematic error caused by optical-axis tilt (Fig. 3d).The X-ray image dataset comprised 100 valid samples (101 were originally acquired, and 1 was discarded because the height reading was missing). After randomization according to the acquisition order (with random rotations of the gimbal platform), the dataset was divided into training, validation, and test sets at a ratio of 8:1:1, corresponding to 80, 10, and 10 samples, respectively. To enable a more comprehensive evaluation of the error distribution under limited-sample conditions, final pose inference was performed on all 100 samples, and the error statistics for each subset were reported independently.

The optical image dataset consisted of 194 valid samples. Each sample included one optical image and the corresponding initial estimate of the relative pose between the camera and the marker. The initial pose was obtained by transforming the physical measurements into a unified checkerboard reference-plane coordinate system, using the measured camera heights (88.5, 101.5, and 112.5 cm) and the

marker planar coordinates (6, 48, 78).Similar to the X-ray dataset, the optical dataset was randomly divided into training, validation, and test sets in an 8:1:1 ratio, resulting in 155, 19, and 20 samples, respectively. The three camera heights were maintained in proportional distribution across the training, validation, and test subsets. Final pose inference was performed on all 194 samples, and the error statistics for each subset were reported independently.

The complete IMU-marker assembly procedures, coordinate system calibration protocols, and data preprocessing details are provided in Supplementary Information Section 3 and 4.

2.4 Evaluation metrics

The evaluation metrics focused on the deviation between the algorithm-predicted pose and the ground-truth pose, and were reported separately by modality and by degree of freedom.

X-ray modality: The basic evaluation metric was defined as the difference between the predicted and measured Euler angles for each axis. Only the X and Y axes were included, while the Z axis was excluded because its IMU measurement was unreliable due to a mechanical malfunction. The errors were further reported separately in the two clinically relevant planes, namely the coronal plane and the sagittal plane.In addition, the deviations along the X and Y axes were combined into an overall angular error, defined as $\mathrm{e}_\theta=\sqrt{(\mathrm{e}_\mathrm{X}^2+\mathrm{e}_\mathrm{Y}^2)}$.which served as a single-value summary of rotational accuracy. This composite metric is approximately equivalent to the geodesic angular error between rotation matrices when the Z-axis component is ignored.

Optical modality: The evaluation metrics were based on the differences between the predicted and ground-truth rotation angles along the Yaw, Pitch, and Roll axes. For each axis, the mean ± standard deviation, median, and maximum error were reported.The overall three-dimensional angular error was defined as $\mathrm{e}_\theta=\mathrm{e}_\mathrm{Y}^2+\mathrm{e}_\mathrm{P}^2+\mathrm{e}_\mathrm{R}^2$ .The detection success rate was defined as the percentage of valid comparison samples among all samples, where a valid sample required successful target detection and at least three detected keypoints.

Pose transfer proof-of-concept:The Euclidean pixel distance between the manually annotated predicted vertices and the ground-truth vertices was used as the metric for point deviation (px). The angle between the predicted edge direction vector and the ground-truth edge direction vector was used as the metric for edge angular deviation(°).Only edges whose both endpoints were annotated were included in the comparison, while edges connected to occluded vertices were excluded.

Reporting format:For all modalities, results were presented in the following forms:(1)Scatter plots of predicted values versus ground-truth values, including the $\mathrm{y}=\mathrm{x}$ reference line and MAE;(2)Histograms of error distributions for each axis or point;(3)Cumulative distribution function (CDF) curves of the composite error;(4)Per-sample or per-edge error plots to facilitate identification of outliers.

Statistical analysis:All pose estimation errors were reported as mean absolute error (MAE) ± standard deviation (SD), with the median and maximum also provided to characterize distributional skewness. The cumulative distribution function (CDF) was used to describe the overall error distribution and corresponding percentiles. The detection success rate was defined as the percentage of valid comparison samples among all samples. For the optical modality, complete pose loss was identified based on the naturally occurring bimodal distribution of the composite error (see Section 3.2 for details). Pearson's correlation coefficient was used for correlation analysis to assess the degree of linear association between two variables. All statistical analyses were implemented in Python 3.10 using NumPy and SciPy, and all related figures were generated in Python.

3. Results

3.1 Evaluation of X-ray Pose Recognition

The accuracy of X-ray pose estimation was evaluated using 100 test samples (IDs 101–200). Owing to a mechanical malfunction of the IMU around the Z-axis (see Section 4.2), the following analysis focused on rotation angles about the X-axis (coronal plane) and Y-axis (sagittal plane) only.

Per-axis accuracy. Figures 4(a–b) show scatter plots of the predicted versus measured rotation angles about the X- and Y-axes. For the X-axis, the mean absolute error (MAE) was 1.584° (SD = 1.256°, median = 1.290°, maximum = 5.505°). For the Y-axis, the MAE was 1.898° (SD = 1.502°, median = 1.560°, maximum = 8.151°).

Error independence. The X–Y error scatter plot (Fig. 4c) showed that the prediction errors along the two axes were nearly uncorrelated (Pearson's r = 0.054).

Overall accuracy. The composite angular error was defined as $e_\theta = \sqrt{(e_X^2 + e_Y^2)}$, with a median of 2.45° and an MAE of 2.73° (SD = 1.58°). The cumulative distribution curve (Fig. 4d) showed that 50% of samples had a composite error < 2.45°, 80% < 3.86°, 90% < 4.82°, and 95% < 6.26°. The per-sample error distribution (Fig. 4e) was generally uniform across the tested pose range.

Error distribution characteristics. The histograms of the error distributions for both axes (Figs. 4f–g) were approximately symmetric and centered around zero (X: mean = 0.02°, SD = 2.01°; Y: mean = − 0.01°, SD = 2.41°).

3.2 Evaluation of optical pose recognition

The evaluation of optical pose estimation was based on 194 optical images. Analysis was performed after target detection and keypoint extraction. Of these, 19 cases (9.8%) were excluded because of target detection failure or because fewer than three keypoints were available. Among the remaining 175 valid comparisons, 4 cases (2.3%; ID = 73, 88, 136, and 159) showed complete pose loss (composite error > 170°). This threshold was determined based on the naturally bimodal error distribution: in 171

samples, the composite error was concentrated within 4°, whereas in these 4 cases, the error was approximately 170°, corresponding to a reflective ambiguity in marker orientation (flipped solution). The 170° cutoff therefore formed a clear natural separation between the normal accuracy range and the failure mode. After excluding these 4 cases, the remaining 171 samples (97.7% of valid comparisons) were regarded as reliable samples.

Per-axis accuracy. The Yaw axis showed the highest accuracy (MAE = 0.252°, SD = 0.161°, median = 0.234°, maximum = 0.774°), followed by the Pitch axis (MAE = 0.411°, SD = 0.273°, median = 0.363°). The Roll axis exhibited the largest error (MAE = 1.438°, SD = 0.930°, median = 1.307°, maximum = 3.905°), which was approximately 5.7 times the MAE of the Yaw axis. Roll-axis data were visualized using broken axes (Figs. 5c1–c3), with the data grouped by angle into three clusters—negative angles, near-zero angles, and positive angles—and displayed separately at enlarged scales to avoid visual distortion caused by the ± 180° wrap-around discontinuity.

Error correlations. The cross-axis scatter plots of the three angular errors (Figs. 5g–i) showed a strong negative correlation between Yaw and Pitch errors (r = − 0.948). Moderate correlations were observed between Yaw and Roll (r = − 0.558) and between Pitch and Roll (r = 0.661).

Overall accuracy. Among the reliable samples, the median composite angular error was 1.38°, and the MAE was 1.52° (SD = 0.98°). The CDF (Fig. 5j) showed that 50% of samples were < 1.38°, 80% < 2.41°, 90% < 2.85°, and 95% < 3.24°. The overall detection success rate was 90.2% (175/194).

Error distribution characteristics. The histograms of the error distributions for all three axes (Figs. 5d–f) were approximately symmetric and centered around zero (Yaw: mean = 0.00°, SD = 0.30°; Pitch: mean = 0.01°, SD = 0.49°; Roll: mean = − 0.01°, SD = 1.72°).

3.3 Proof-of-concept for marker-to-object pose transfer

To visually demonstrate the transferability of the estimated pose from the marker to the target object, this section presents a proof-of-concept using a regular geometric object (a wooden block). Owing to the limited camera field of view, which allowed only a narrow range of viewpoints in which both the marker and the target object could be fully captured, four representative optical images were selected for fitting demonstration, covering different viewing angles (approximately 0°–135° in azimuth; Figs. 6a–c). In each image, the seven visible vertices of the wooden block were annotated with the algorithm-predicted positions (red dots) and the ground-truth positions (green dots). After excluding vertices with missing annotations (vertex 5 in image 2; vertices 3, 5, and 7 in image 4), a total of 24 point-pair comparisons and 28 edge-pair comparisons were obtained across the four images.

Point deviation (Fig. 6d). The mean Euclidean distance between the predicted and ground-truth vertices was 26.6 px (SD = 14.9 px, median = 25.7 px, range = 3.6–58.4 px). Vertices farther from the camera—that is, those with greater depth relative to the nearest vertex (vertex 4)—showed systematically larger errors: the mean errors for near-camera vertices (2, 4, and 6) ranged from 16.8 to 20.6 px, whereas those

for far-camera vertices (1, 3, 5, and 7) ranged from 29.0 to 43.4 px. Point 7 had the largest cross-image mean error (43.4 px, N = 3), while point 2 had the smallest mean error (16.8 px, N = 4).

Edge angular deviation (Fig. 6e). For each image, the angle between the predicted edge direction vector and the ground-truth edge direction vector was calculated for edges whose two endpoints were both annotated. A total of nine edges were included, excluding the three edges connected to the occluded vertex 8. The mean edge angular deviation was 3.10° (SD = 3.09°, median = 2.32°, range = 0.00–11.97°). The three edges connected to the nearest vertex 4 (4–2, 4–5, and 4–6) showed relatively small angular errors, with a mean of 2.45°, whereas edges connected to vertices farther from the camera exhibited larger angular deviations, consistent with the trend observed for point deviations. The largest edge angular deviation (11.97°, edge 5–7 in image 3) occurred between two far-camera vertices, both of which also had relatively large point deviations.

Point–edge correlation. The scatter plot of mean point error versus edge angular deviation (Fig. 6f) showed an approximately zero correlation between the two metrics (r = 0.078).

Across the four images covering different viewing angles, the error patterns were consistent and in agreement with theoretical expectations, confirming that the estimated pose can be stably transferred from the marker to the target object. This provides a foundation for subsequent quantitative evaluation in real surgical scenarios.

3.4 Cross-modal registration consistency

A prerequisite for using the marker as a pose-transfer medium is that the two recognition modalities provide comparable accuracy and share a consistent coordinate basis. It should be noted that the pose ground truth for the X-ray and optical modalities was obtained from different measurement systems (see Section 4.5). Therefore, the following comparison reflects the performance of each modality relative to its own reference standard, rather than a direct accuracy comparison under a single unified coordinate system. To verify this prerequisite, the pose estimation accuracy of the X-ray and optical modalities was systematically compared (Fig. 7).

Axis-wise comparison. Optical yaw (MAE = 0.252°) and pitch (MAE = 0.411°) were markedly more accurate than X-ray X-axis rotation (1.584°) and Y-axis rotation (1.898°), amounting to only 15.9% and 21.7% of the corresponding X-ray-axis MAEs, respectively. Optical roll (MAE = 1.438°) was of the same order of magnitude as X-ray Y-axis rotation (1.898°), corresponding to approximately 75.8% of the latter. Within each modality, the ranking of precision was consistent: the axis aligned with the marker's dominant geometric features showed the highest accuracy, whereas the axis corresponding to in-plane rotation showed the lowest accuracy.

Overall error comparison. The overlaid CDFs (Fig. 7b) showed that the optical modality achieved better overall accuracy than the X-ray modality and exhibited a more concentrated error distribution. The median composite error decreased from 2.45° (X-ray) to 1.38° (optical), a reduction of 43.7%, while the standard deviation decreased from 1.58° to 0.98°, corresponding to a 38.0% reduction.

Dispersion analysis. The box plots (Fig. 7c) showed that the optical yaw axis had the most concentrated error distribution (IQR = 0.22°), which was only 19.0% of the IQR of the X-ray X axis (1.16°). The X-ray Y axis showed the greatest dispersion (IQR = 2.19°), with tail outliers corresponding to extreme sagittal-plane tilts greater than 40°, consistent with the robustness analysis in Section 2.1. The dispersion of the optical roll axis (IQR = 1.36°) was also markedly smaller than that of both X-ray axes.

4. Discussion

4.1 Main findings

This study is the first to systematically derive three design constraints for surgical navigation markers from the physical constraints of MR head-mounted displays: omnidirectional recognizability, X-ray recognizability, and multimodal registration. Based on these constraints, we designed a marker configuration consisting of a cube combined with four cylindrical axes of unequal diameters.The experimental results support the design hypothesis at three levels. Firstly, a single marker configuration successfully enabled bimodal pose recognition. In the X-ray modality, accuracy remained consistent between the coronal and sagittal planes, while in the optical modality, angular errors were concentrated within 2.5° (CDF80%). Together, these findings confirm the feasibility of pose recognition in both modalities within their respective sensing channels.Secondly, the unequal-diameter cylindrical axes could be stably distinguished by width ranking in X-ray projections, validating the theoretical expectation that the unequal-diameter design resolves correspondence ambiguity. Thirdly, in the pose-transfer validation experiment, vertex errors showed a systematic gradient along the depth direction, whereas edge-angle deviations remained around 3° and were not correlated with point-wise errors. This indicates that the pose estimation preserved structural orientation fidelity despite depth-dependent positional deviations.

The consistency of the cross-modal error structure further supports these conclusions. In both modalities, the ranking of accuracy was identical: the axis aligned with the marker's primary geometric features achieved the highest accuracy, whereas the in-plane rotational axis showed the lowest accuracy.The spatial distribution of errors in the X-ray results is consistent with the robustness analysis in Appendix 1.6. The largest overall error (ID 200, 9.6°) occurred under an extreme sagittal-plane angle condition $(\mathrm{Y}_{\mathrm{real}}>40^{\circ})$, in which one cylindrical axis became nearly aligned with the X-ray beam direction. Under this condition, the available projection information was effectively reduced to that of three cylinders, leading to increased—but not divergent—error, in agreement with the theoretical prediction of a "degraded but solvable" configuration.Similarly, the four cases of complete pose loss in the optical modality (overall error $>170^{\circ}$) also occurred under extreme viewing angles, consistent with the degeneracy analysis in the appendix. All of these cases could be reliably detected and flagged as failures using an overall error threshold, and therefore do not affect accuracy under normal operating conditions.

The error distributions in both modalities exhibited symmetric, zero-centered characteristics (mean error for each axis < 0.02° in X-ray and < 0.01° in the optical modality), ruling out systematic bias and indicating that the residual errors were dominated primarily by random noise.However, the correlation structure of the errors differed markedly between the two modalities. In the X-ray modality, errors along the two axes were nearly uncorrelated (Pearson $\mathrm{r}=0.054$), and no coupled failure pattern involving simultaneous degradation of both axes was observed under specific poses. In contrast, the optical modality exhibited a distinctly different pattern: yaw and pitch errors were strongly negatively correlated ($\mathrm{r}=-0.948$), while roll showed moderate correlations with both ($|\mathrm{r}|=0.56\backslash\text{--}0.66$).This contrast suggests that the cylindrical projection width in the X-ray modality provides additional information that helps the algorithm discriminate between rotation directions, whereas this information is absent in the purely optical modality.

The proof-of-concept pose-transfer experiment further revealed a practically meaningful property: vertex pixel deviation was almost uncorrelated with edge-angle deviation ($\mathrm{r}=0.078$), indicating that pose estimation errors are decoupled across spatial scales. Specifically, vertex displacement tends to manifest as a global translation rather than as independent perturbations, thus preserving the structural orientation of the edges. This suggests that, under the experimental test conditions, the marker provides more stable angular information than positional information.

Taken together, the single-marker configuration could be stably identified in both optical and X-ray images and yielded comparable pose parameters (overall MAE: 1.52° for optical vs. $2.73^\circ$for X-ray), thus validating the core hypothesis of Design Constraint 3(multimodal registration).The differences between the two modalities in terms of absolute accuracy and error structure should not be regarded as a flaw of the design. Rather, they can be further improved through refinements in the data acquisition protocol—such as replacing the IMU with external optical tracking as the ground-truth reference for X-ray pose—and through targeted algorithmic optimization aimed at improving pitch–yaw decoupling.

4.2 Analysis of Error Sources

The accuracy differences between the X-ray and optical modalities can be discussed from three levels.

Error sources in the X-ray modality: The resolution of X-ray imaging is limited by the detector pixel size (approximately 0.3 mm/pixel). Under the condition of SDD = 100 cm, the localization uncertainty of a single pixel already corresponds to a non-negligible angular deviation.In addition, the IMU sensor exhibited drift along the Z axis due to interference from surrounding metallic equipment, rendering the Z-axis angle annotations unreliable; accordingly, these data were excluded from the present study.The Z-axis data used in the subsequent fitting were obtained by jointly recovering the full rotation matrix from the cylindrical constraints in the X-ray projection and the known X- and Y-axis Euler angles, from which the Z angle was then extracted; the specific algorithmic procedure is described in Section 2.1. In this study, the algorithmically recovered Z angle was used only in an auxiliary manner and was not treated as an independent evaluation metric. The subpixel accuracy of cylindrical edge extraction is a key factor determining the lower bound of pose estimation error.

Furthermore, the Y-axis error in the X-ray modality exhibited mild heteroscedasticity (r = 0.41 between |error| and |angle|), indicating that the greater the marker tilt, the larger the error tended to be. This is consistent with the increased uncertainty in edge extraction from cylindrical projections under conditions of extreme perspective foreshortening.However, the two-axis errors in the X-ray modality were almost uncorrelated (Pearson r = 0.054), and no coupled failure pattern involving simultaneous degradation of both axes was observed for any specific pose. This contrasts with the strong Yaw–Pitch coupling seen in the optical modality discussed below, suggesting that the error sources in the two modalities are different in nature.

Error sources in the optical modality: physical measurements (measuring tape and plumb-line positioning) used to establish the pose ground truth introduced uncertainties on the order of several millimeters, thus limiting the lower bound of optical error evaluation. Residual errors in the flatness of the checkerboard reference plane and in the alignment of the camera optical axis likewise contributed to systematic bias.

Among the 194 samples, complete pose loss (composite error > 170°) occurred in 4 cases (2.3%) under extreme viewing conditions, when one cylindrical axis became nearly parallel to the optical axis, leading to target detection mismatches. This failure mode is consistent with the prediction in Appendix 1.6, and all 4 cases could be reliably detected and flagged as failures using a composite error threshold, without affecting accuracy under normal operating conditions.

Error correlation structure in the optical modality contrasted sharply with that in the X-ray modality: Yaw and Pitch errors exhibited a strong negative correlation (r = − 0.948), indicating a structured directional trade-off of the algorithm in the pitch–yaw plane.This coupling may arise because, under single-view conditions, the projection changes of the cube corners along the pitch and yaw directions are partially similar, limiting the ability to distinguish between the two from a single optical image. Notably, no comparable coupling was observed in the X-ray modality (r = 0.054), possibly because the projected width of the cylinders helps the X-ray algorithm discriminate rotation direction, whereas the optical algorithm, relying only on corner and surface features, has greater difficulty separating the two.

The two modalities showed complementary error characteristics: the X-ray modality was stronger in rotational constraint, as the orientations of the four cylindrical axes provided redundant directional information, whereas the optical modality was stronger in translational constraint, since the image coordinates of the cube corners provided direct planar positional information.This complementarity directly validates the underlying logic of multimodal registration within the three design constraints: different modalities leverage their respective geometric strengths to jointly establish a complete six-degree-of-freedom registration pathway.

Across the two modalities, the optical modality achieved higher accuracy than the X-ray modality (overall MAE: 1.52° vs 2.73°), and this difference may be attributed to two factors:(1) The initial optical pose estimate was refined through a PnP procedure based on the marker cube corners. Specifically, Levenberg–Marquardt optimization was used to minimize the reprojection error. In contrast, the X-ray

pose ground truth was obtained directly from IMU readings and was not subjected to comparable post-processing correction.(2) A mechanical failure in the IMU z-axis introduced additional uncertainty into the ground-truth annotation of the X-ray sagittal plane (y-axis). As a result, the annotation accuracy of the X-ray modality was lower than that of the optical modality. Therefore, at least part of the performance gap between the two modalities may be attributable to unequal annotation quality. Whether this level of accuracy is sufficient for specific surgical procedures still requires validation against the relevant clinical accuracy standards in the intended application scenario.

4.3 Comparison with Existing Methods

For X-ray markers, Hatt et al. designed a marker device based on the spatial distribution of spheres, which is capable of identifying pose from a single X-ray image.[21]We further took into account the omnidirectional requirements of the C-arm and incorporated optical compatibility.Grupp et al. proposed a pose estimation method based on intraoperatively implanted bone markers during PAO.[22]In contrast, the marker proposed in this study is connected to the bone via Kirschner wires, resulting in smaller incisions, simpler installation and removal, and broader applicability, while also enabling pose recognition by an optical probe.

In terms of optical markers, the tetrahedral infrared marker proposed by Immoor and Koller and the cubic marker proposed by Sanchez-Sotelo et al. both addressed the recognition performance and omnidirectional detectability of non-spherical markers. However, these designs are limited to infrared optical tracking and do not possess X-ray imaging capability.[11, 19]Zach and Stelldinger (2025) explored how to design surface patterns on cylinders to enable pose recognition with visible-light cameras, further providing a recognition solution for non-spherical markers.[20]The conventional commercial NDI Polaris optical tracking system has a nominal translational accuracy of approximately 0.25 mm RMS. However, it relies on an externally fixed tracker, provides optical tracking only without X-ray recognition capability, and its marker base suffers from optical blind spots.[14, 15]Although the optical accuracy of the marker proposed in this study has not yet reached the level of the NDI Polaris system, the equivalent displacement error is approximately 1.1 mm when calculated at the point on the marker surface farthest from the center of rotation (about 4 cm away). Nevertheless, it provides X-ray-based pose updating capability and omnidirectional detectability, both of which are unavailable in the NDI system and are core requirements for MR HMD–based navigation scenarios.Kunz et al.'s systematic evaluation of various medical markers also pointed out that existing marker schemes do not simultaneously satisfy the combined requirements of multimodal imaging and omnidirectional tracking. The design proposed in this study represents a first response to this gap.[16]

At the level of MR surgical navigation systems, Muller et al. reported the use of AR navigation based on intraoperative 3D imaging for pedicle screw placement, while Sanchez-Sotelo et al. validated the accuracy of mixed-reality surgical navigation in shoulder arthroplasty.However, both navigation systems relied on optical tracking and therefore still required manual registration by the surgeon.[10, 11]In contrast, the pose-transmission chain proposed in this study—where X-ray imaging acquires the relative pose between the marker and the bone, optical tracking acquires the relative pose between the HMD and

the marker, and the HMD thus indirectly obtains the bone pose—provides an intraoperative registration-update strategy for MR navigation using the marker as an intermediary, thus avoiding the limitations of relying solely on a one-time preoperative registration.In future work, we will consider conducting accuracy benchmarking under the same experimental conditions against commercial HMD-based navigation systems (such as Augmedics xVision) and conventional optical tracking systems (such as NDI Polaris). This has already been included in our planned collaborative research.

Across these comparisons, in this study, our marker design inherits both the geometric paradigm of recovering three-dimensional pose from two-dimensional X-ray marker images and the omnidirectional recognition concept of optical markers.By combining the two, we establish a dual-modal pose-transmission chain from X-ray to optical tracking. This enables precise coordination between the MR device and the medical imaging system. In turn, it supports the purpose of surgical navigation.Compared with existing solutions that serve only a single modality, our marker offers dual-modal registration capability and broader adaptability across surgical navigation scenarios.

4.4 Limitations

Our study has the following limitations, which should be taken into consideration when interpreting the results.(1) Scope of geometric configuration validation:This study only validated the design logic of the specific configuration of a “cube plus four cylinders with unequal diameters.” Whether manufacturing complexity can be further reduced and accuracy improved by decreasing the number of cylinders (e.g., to a three-axis design), altering the type of geometric primitives (e.g., replacing cylinders with prisms), or optimizing dimensional proportions remains unexplored.(2) Data scale and diversity:The X-ray dataset comprised 100 samples in total (training/validation/test = 80/10/10), and the optical dataset comprised 194 samples (155/19/20), which remain small-sample datasets for deep learning models.To ensure the robustness of the error statistics, inference was performed on all samples, although the accuracy on training-set samples may be slightly overestimated. While pose coverage in the dataset was randomized, it was not systematically stratified, and performance degradation under certain extreme viewing angles may therefore have been overlooked. Larger and more diverse datasets would facilitate a more reliable evaluation of generalization performance.(3) Limitations in model morphology:This study only validated the fitting approach using regular geometric models and did not account for the irregular morphology of real bones, the effects of density differences between cortical and cancellous bone on X-ray projection, or interference from surrounding soft tissues. Substantial intermediate validation will still be required before translation from the model setting to clinical application.(4) Limitations in the reference standard for cross-modal accuracy comparison:The ground-truth pose in the X-ray modality was obtained from IMU sensor readings without post-processing correction, while the ground-truth pose in the optical modality was based on physical measurements relative to a checkerboard reference plane (using a tape measure and plumb-line positioning). Both reference standards therefore involve baseline uncertainties on the order of several millimeters and several degrees.This limitation implies that the cross-modal accuracy differences reported in this study should be interpreted as deviations of each modality relative to its own reference standard, rather than as a direct absolute comparison between the two modalities.

Future studies should introduce a unified external optical tracking system, such as NDI Polaris, as a common reference standard for both modalities, in order to eliminate the influence of differing baselines on accuracy evaluation.(5) Lack of direct comparison with conventional systems:Due to limitations in equipment accessibility, this study did not perform a head-to-head accuracy comparison with commonly used commercial surgical navigation systems under the same experimental conditions. In future work, benchmarking against commercial HMD-based navigation systems, such as Augmedics xVision, as well as conventional optical tracking systems, such as NDI Polaris, should be considered. Such comparative experiments have already been included in our planned collaborative follow-up studies.(6) Limitations in the X-ray ground-truth reference:The ground-truth pose for the X-ray modality was derived from a consumer-grade IMU sensor, whose bias and drift may introduce additional uncertainty into the reference itself. As a result, the X-ray accuracy reported in this study is likely to be conservative, and the actual performance of the algorithm may be better than indicated.In addition, the yaw angle about the Z-axis was considered unreliable due to sensor drift, resulting in the absence of one rotational degree of freedom in the X-ray pose evaluation. Although Z-axis information could be indirectly estimated through image-based recognition and formula-based conversion, a complete three-dimensional accuracy validation still requires an alternative ground-truth reference, such as an external optical tracking system (e.g., NDI Polaris) or a robotic arm. This experiment has already been incorporated into our planned future work.(7) Absence of a standardized benchmark for accuracy evaluation:In this study, the accuracy of the X-ray modality was evaluated against a consumer-grade IMU, while the optical modality was referenced to physical measurements of a checkerboard. Neither of these serves as a widely recognized standard reference in the field of surgical navigation, such as the NDI Polaris optical tracking system.This limitation not only reduces the direct comparability of the absolute accuracy values reported in this study with results from other literature based on standardized reference systems, but also makes it difficult to disentangle algorithmic error from benchmark-related error in the observed cross-modality discrepancies. To address this issue, our team has incorporated a unified dual-reference framework into the design of future experiments, using both external optical tracking and robotic-arm kinematics as common benchmarks.

5. Conclusion

This study focused on the problem of marker design for mixed-reality surgical navigation and proposed a mechanistic design framework based on three constraints: omnidirectional recognizability, X-ray recognizability, and multimodal registration. Guided by this framework, a novel marker configuration was designed and experimentally validated.Experimental results showed that the proposed marker achieved effective pose recognition under both X-ray and optical modalities, providing preliminary validation of the feasibility of a constraint-driven marker development strategy. These findings also offer foundational support for a marker-mediated pose transfer chain in mixed-reality surgical navigation.

The core contribution of this work lies not in a single accuracy metric, but in the establishment of a generalizable methodology: starting from the physical constraints of the target system, deriving the essential geometric properties required of the marker, and then achieving pose estimation through a

combination of geometric solvability and data-driven methods.This framework may also serve as a useful reference for marker design in other surgical navigation scenarios, where it can be adapted by adjusting the design constraints according to specific sensing conditions and procedural requirements. As the imaging capability and computational power of MR head-mounted displays continue to improve, and as intraoperative imaging systems such as C-arm, O-arm, and cone-beam CT become increasingly widespread, automatic X-ray/optical multimodal registration is expected to emerge as a standard configuration in surgical navigation. The marker design and pose estimation method presented in this study provide a preliminarily validated technical approach toward that goal.

Declarations

Acknowledgments

The authors thank the staff of the Department of Radiology,Xiangya Hospital, for their assistance with X-ray image acquisition. We also thank Professor Shenghui Liao and his team at the School of Computer Science and Engineering, Central South University, for their guidance and contributions to the artificial intelligence algorithm design. We are grateful to Jinse Medical Technology Co.,Ltd. for providing technical support during the early exploratory stage of this work; their QR-code recognition software developed for Apple Vision Pro served as a valuable preliminary investigation for this study.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 82472609); the Hunan Provincial Research Project on Degree and Graduate Education Reform (Grant No.2025JGYB039); and the Beijing Association for Chronic Disease Prevention and Treatment of Integrated Traditional Chinese and Western Medicine Horizontal Collaborative Research Project (Contract No. 2025MB8819).

Conflict of interest

This work was supported in part by a horizontal collaborative research project (Contract No. 2025MB8819) with the Beijing Association for Chronic Disease Prevention and Treatment of Integrated Traditional Chinese and Western Medicine. The sponsor had no role in the study design, data collection, analysis, manuscript preparation, or the decision to submit for publication. The authors declare no other competing interests.

Ethical statement

This study did not involve human or animal subjects.

Data availability

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Author contributions

G.Z.X. conceived the marker design, designed the experiments, performed the theoretical derivations, conducted the experimental work, and wrote the original manuscript. X.L. contributed to the literature review, participated in the experimental design discussions, and assisted in the experiments. W.Y.X. developed the pose estimation algorithms and processed the experimental data. Z.Y. and Y.X.Y. assisted in the experiments. X.P.W. constructed the digital model of the marker and performed the 3D printing. D.Z., H.L., Y.X., and K.L.C. provided guidance on the manuscript and contributed to its review and editing. C.G.W. conceived the overarching concept, acquired the funding, and supervised the entire project. All authors reviewed and approved the final manuscript.

References

1. Sloan, M., Premkumar, A. & Sheth, N. P. Projected Volume of Primary Total Joint Arthroplasty in the U.S., 2014 to 2030. *J. Bone Joint Surg. Am.* **100**, 1455–1460 (2018).
2. Schwartz, A. M., Farley, K. X., Guild, G. N. & Bradbury, T. L. Projections and Epidemiology of Revision Hip and Knee Arthroplasty in the United States to 2030. *J. Arthroplasty*. **35**, S79–S85 (2020).
3. Kurtz, S., Ong, K., Lau, E., Mowat, F. & Halpern, M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J. Bone Joint Surg. Am.* **89**, 780–785 (2007).
4. Singh, J. A., Yu, S., Chen, L. & Cleveland, J. D. Rates of Total Joint Replacement in the United States: Future Projections to 2020–2040 Using the National Inpatient Sample. *J. Rheumatol.* **46**, 1134–1140 (2019).
5. World Health Organization. *Osteoarthritis Fact Sheet* (World Health Organization, 2023).
6. Kosmopoulos, V. & Schizas, C. Pedicle screw placement accuracy: a meta-analysis. *Spine* **32**, E111–120 (2007).
7. Sadhu, A. et al. Acetabular Component Position and the Risk of Dislocation Following Primary and Revision Total Hip Arthroplasty: A Matched Cohort Analysis. *J. Arthroplasty*. **32**, 987–991 (2017).
8. Sang, A. Y., Wang, X. & Paxton, L. Technological Advancements in Augmented, Mixed, and Virtual Reality Technologies for Surgery: A Systematic Review. *Cureus* **16**, e76428 (2024).
9. Móga, K., Hölgyesi, Á., Zrubka, Z., Péntek, M. & Haidegger, T. Augmented or Mixed Reality Enhanced Head-Mounted Display Navigation for In Vivo Spine Surgery: A Systematic Review of Clinical Outcomes. *J. Clin. Med.* **12**, 3788 (2023).
10. Müller, F. et al. Augmented reality navigation for spinal pedicle screw instrumentation using intraoperative 3D imaging. *Spine J.* **20**, 621–628 (2020).
11. Sanchez-Sotelo, J. et al. Validation of mixed-reality surgical navigation for glenoid axis pin placement in shoulder arthroplasty using a cadaveric model. *J. Shoulder Elb. Surg.* **33**, 1177–1184 (2024).
12. Bian, D. et al. The application of extended reality technology-assisted intraoperative navigation in orthopedic surgery. *Front. Surg.* **11**, 1336703 (2024).

13. Farshad, M., Fürnstahl, P. & Spirig, J. M. First in man in-situ augmented reality pedicle screw navigation. *North. Am. Spine Soc. J. NASSJ*. **6**, 100065 (2021).
14. Fitzpatrick, J. M., West, J. B. & Maurer, C. R. Predicting error in rigid-body point-based registration. *IEEE Trans. Med. Imaging*. **17**, 694–702 (1998).
15. West, J. B. & Maurer, C. R. Designing optically tracked instruments for image-guided surgery. *IEEE Trans. Med. Imaging*. **23**, 533–545 (2004).
16. Kunz, C., Genten, V., Meissner, P. & Hein, B. Metric-based evaluation of fiducial markers for medical procedures. in *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling* (eds Fei, B. & Linte, C. A.) 97SPIE, San Diego, United States, (2019). 10.1117/12.2511720
17. Tsui, D., Jo, M., Nguyen, B., Ahadian, F. & Talke, F. E. Optical Surgical Navigation: A Promising Low-cost Alternative. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.* 1–4 (2023). (2023).
18. Lee, S. et al. A Skin-Conformal, Stretchable, and Breathable Fiducial Marker Patch for Surgical Navigation Systems. *Micromachines* **11**, 194 (2020).
19. Immoor, E. & Koller, T. Potential of Tetrahedral Markers for Infrared Pose Tracking in Surgical Navigation. *Eurographics Workshop Vis. Comput. Biol. Med.* https://doi.org/10.2312/VCBM.20241184 (2024).
20. Zach, J., Stelldinger, P. & PuzzlePoles Cylindrical Fiducial Markers Based on the PuzzleBoard Pattern. Preprint at (2025). https://doi.org/10.48550/ARXIV.2511.19448
21. Hatt, C. R., Raval, A. N. & Speidel, M. A. (eds) A Fiducial Apparatus for 6DOF Pose Estimation of an External Echo Probe from a Single X-ray Projection: Initial Simulation Studies on Design Requirements. in *World Congress on Medical Physics and Biomedical Engineering, June 7–12*, Toronto, Canada (ed. Jaffray, D. A.) 839–842 (Springer International Publishing, Cham, 2015). (2015). 10.1007/978-3-319-19387-8_205
22. Grupp, R. B. et al. Fast and automatic periacetabular osteotomy fragment pose estimation using intraoperatively implanted fiducials and single-view fluoroscopy. *Phys. Med. Biol.* **65**, 245019 (2020).
23. Shiu, Y. C. & Zhuang, H. Pose determination of cylinders, cones, and spheres from perspective projections. in (ed. Bowyer, K. W.) 771–782Orlando, FL, (1992). 10.1117/12.58624
24. Doignon, C. & de Mathelin, M. A. Degenerate Conic-Based Method for a Direct Fitting and 3-D Pose of Cylinders with a Single Perspective View. in *Proceedings IEEE International Conference on Robotics and Automation* 4220–4225 (2007). 4220–4225 (2007). (2007). 10.1109/ROBOT.2007.364128
25. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (IEEE, Las Vegas, NV, USA, 2016). 770–778 (IEEE, Las Vegas, NV, USA, 2016). (2016). 10.1109/CVPR.2016.90
26. Ruiz, N., Chong, E. & Rehg, J. M. Fine-Grained Head Pose Estimation Without Keypoints. in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* 2155–215509 (IEEE, Salt Lake City, UT, USA, 2018). 2155–215509 (IEEE, Salt Lake City, UT, USA, 2018). (2018). 10.1109/CVPRW.2018.00281

27. Ronneberger, O., Fischer, P., Brox, T. & U-Net Convolutional Networks for Biomedical Image Segmentation. in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 (eds (eds Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) 234–241 (Springer International Publishing, Cham, doi:10.1007/978-3-319-24574-4_28. (2015).

Figures

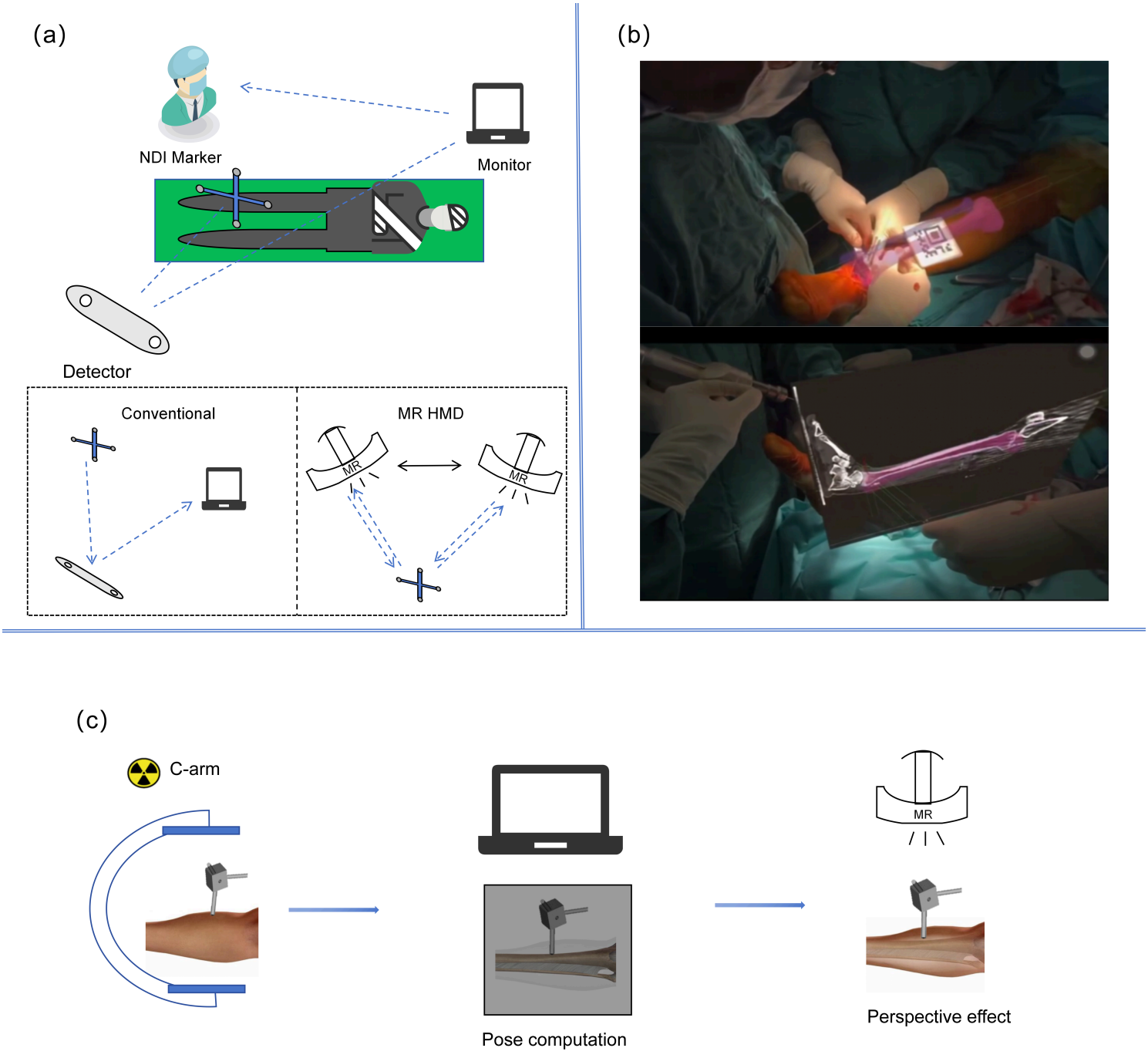


Figure 1

Conceptual overview: from conventional navigation to the proposed approach. (a) In conventional surgical navigation, spatial information is presented on an external 2D monitor, forcing the surgeon to repeatedly shift gaze between the surgical field and the display. Although MR head-mounted displays (HMDs) address this gaze-diversion problem, the tracking camera is now fixed to the surgeon's head and moves continuously, causing the line-of-sight convergence zone to overlap with the operating area. (b) In our preliminary work, in collaboration with Jinse Medical Technology Co., Ltd.,we combined an MR HMD with QR-code fiducial markers to overlay virtual bone models onto preoperative CT images. However, QR-code detection was reliable only from near-frontal viewing angles, and no intraoperative X-ray-based pose update was available, limiting registration accuracy. (c) The proposed dual-modal marker: under X-ray, the four unequally sized cylinders enable marker pose estimation via projection width analysis; combined with existing X-ray-based bone recognition techniques, the marker-to-bone relative pose can be determined. Under optical (HMD) imaging, the cube corner features enable marker pose estimation, establishing the HMD-to-marker spatial relationship. Through the marker as a rigid intermediary, the bone pose is transmitted to the MR view—enabling see-through visualization. This study focuses on validating the feasibility of dual-modal marker pose recognition; bone recognition and relative pose transfer are beyond the current experimental scope.

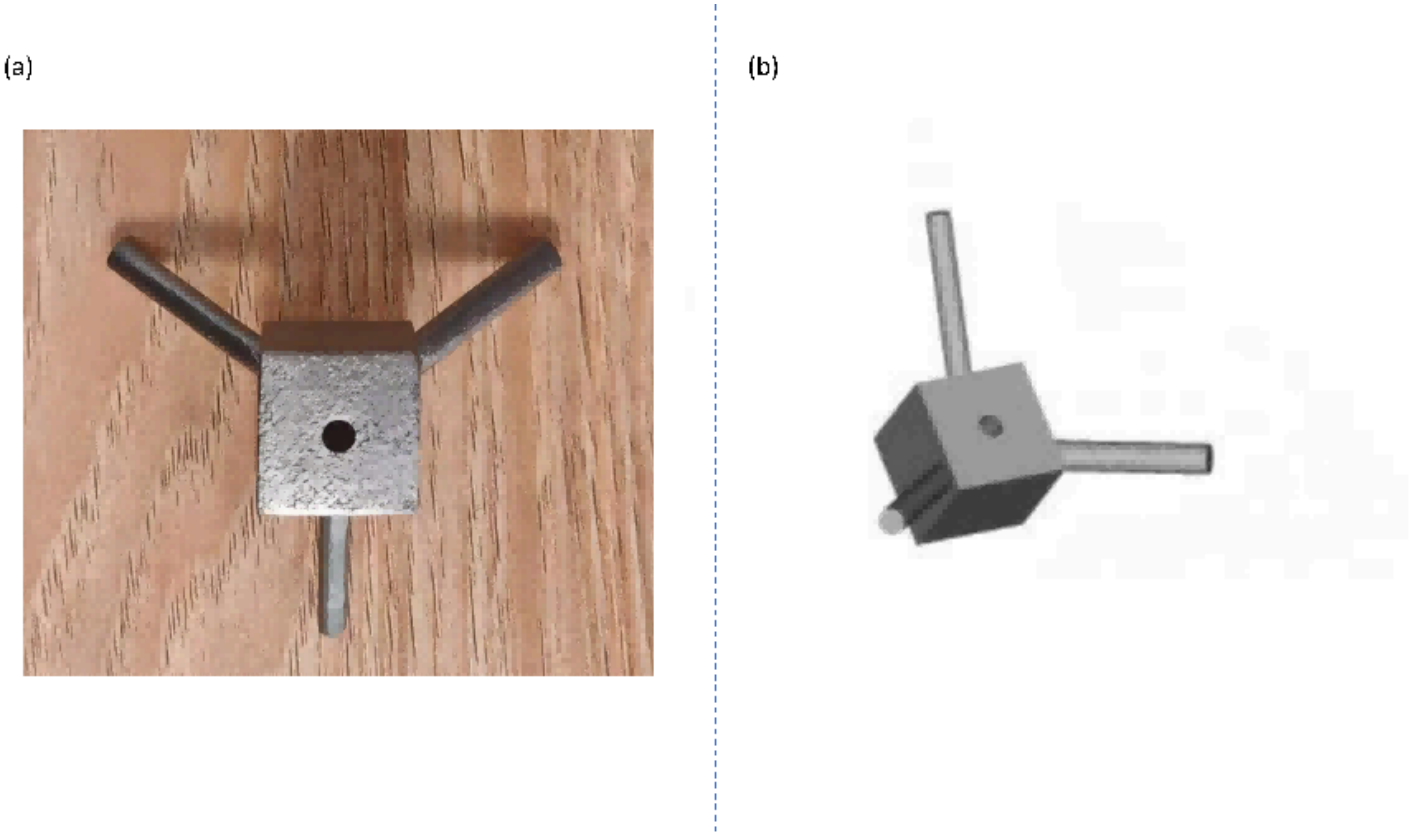


Figure 2

(a) Photograph of the fabricated marker, showing the central cube (providing multi-face optical features for omnidirectional recognition) and four cylindrical arms with unequal diameters ($r_1 < r_2 < r_3 < r_4$)

extending toward the vertices of a regular tetrahedron. (b) CAD rendering with the tetrahedral direction vectors and cylinder radii labeled.

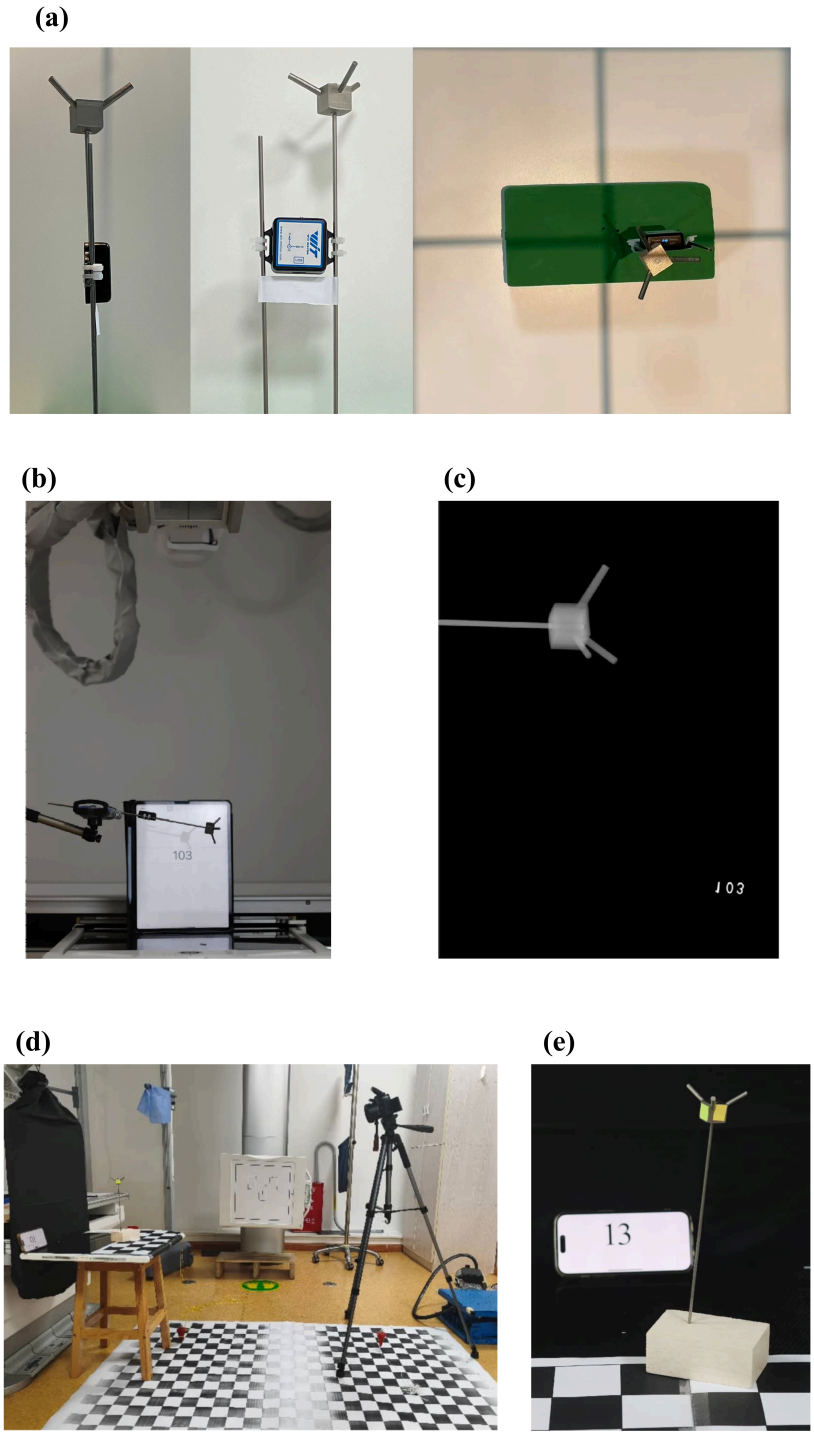


Figure 3

(a) Assembled IMU-marker assembly (front, side, and top views). The IMU sensor (BWT901C) is rigidly attached to the marker via Kirschner wires secured with nylon cable ties and adhesive. (b) X-ray data

acquisition setup (GE Definium 6000, 120 kV, 400 mA, source-to-detector distance = 100 cm). The IMU-marker assembly is mounted on a universal rotation stage between the X-ray source and flat-panel detector. (c) A representative X-ray fluoroscopy image showing four distinguishable cylinder projections (r_1 to r_4). Lead numbers on the background plate provide image indexing for data alignment. (d) Optical data acquisition setup. A Canon R7 camera is mounted on a tripod above a checkerboard reference plane (6 cm × 6 cm grid). The marker is placed on the checkerboard at known coordinates. (e) A representative optical image. The cube faces and cylinder tips provide keypoint features for PnP-based pose refinement.

Fig. 4: X-ray Pose Estimation Results (N = 100, Samples 101-200)

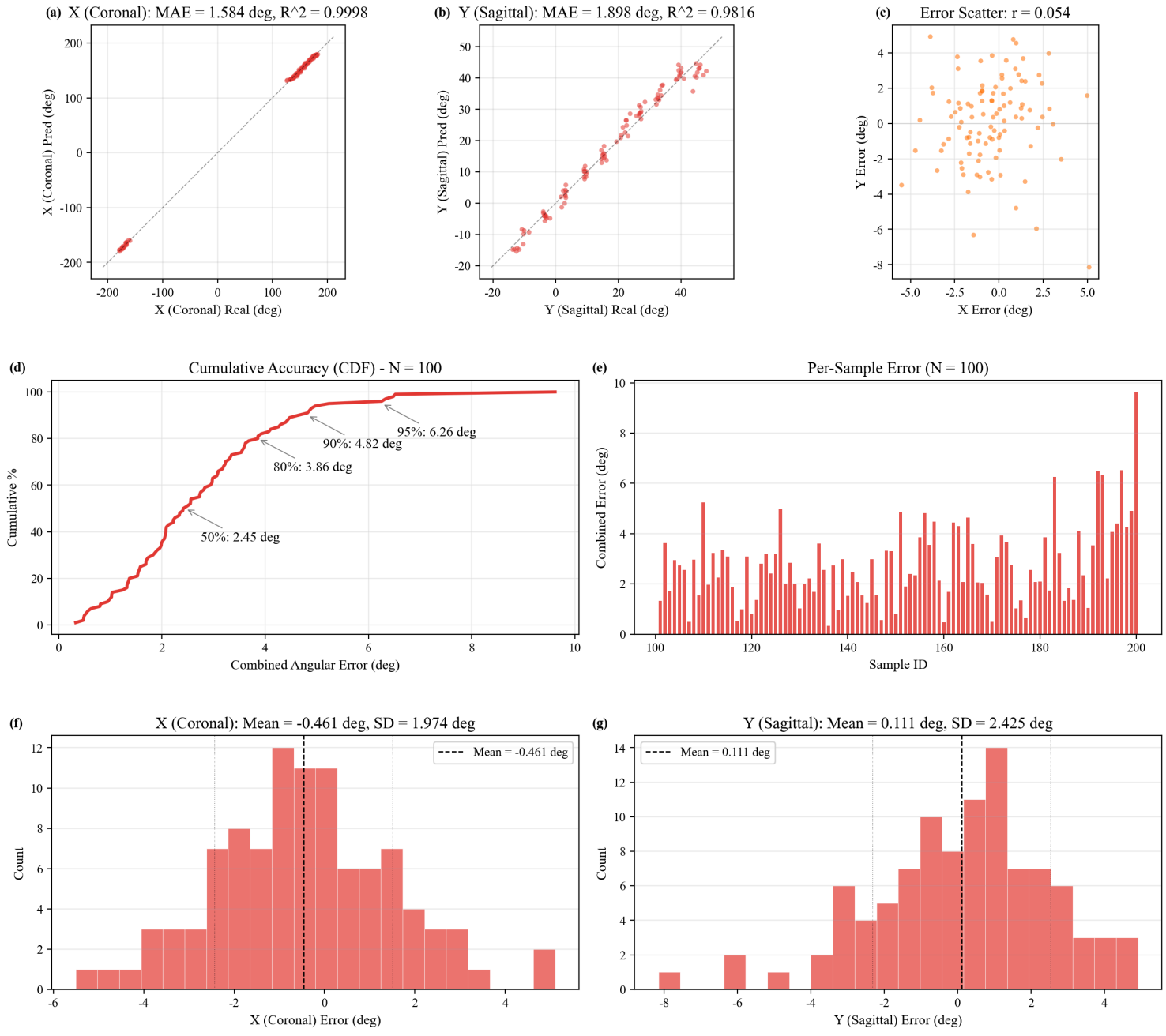


Figure 4

X-ray pose estimation results.

(a) Predicted vs. real X-axis (coronal) rotation angle.MAE = 1.584 deg, SD = 1.256 deg, median = 1.290 deg, max = 5.505 deg. Dashed line: y = x. R^2 = 0.692.**(b) Predicted vs. real Y-axis (sagittal) rotation angle.**MAE = 1.898 deg, SD = 1.502 deg, median = 1.560 deg, max = 8.151 deg. Y-axis accuracy is slightly lower than X-axis with mild heteroscedasticity (|error| correlated with |angle|, r = 0.41).**(c) X-Y error scatter plot.**Near-zero cross-correlation (Pearson r = 0.054); the two axes errors are independent. No coupled failure mode observed.**(d) CDF of combined angular error e_theta = sqrt(e_X^2 + e_Y^2).**N = 100. 50th percentile = 2.45 deg, 80th = 3.86 deg, 90th = 4.82 deg, 95th = 6.26 deg. Median = 2.45 deg.**(e) Combined angular error per sample (ID 101-200).**Error magnitude is broadly uniform across the tested pose range. Largest error (ID 200, combined = 9.6 deg) corresponds to extreme sagittal angle (Y_real > 40 deg).**(f) X-axis (coronal) error distribution.**Mean = 0.02 deg, SD = 2.01 deg. Approximately symmetric and zero-centered; no systematic bias.**(g) Y-axis (sagittal) error distribution.**Mean = -0.01 deg, SD = 2.41 deg. Distribution consistent with X-axis; random noise dominates.

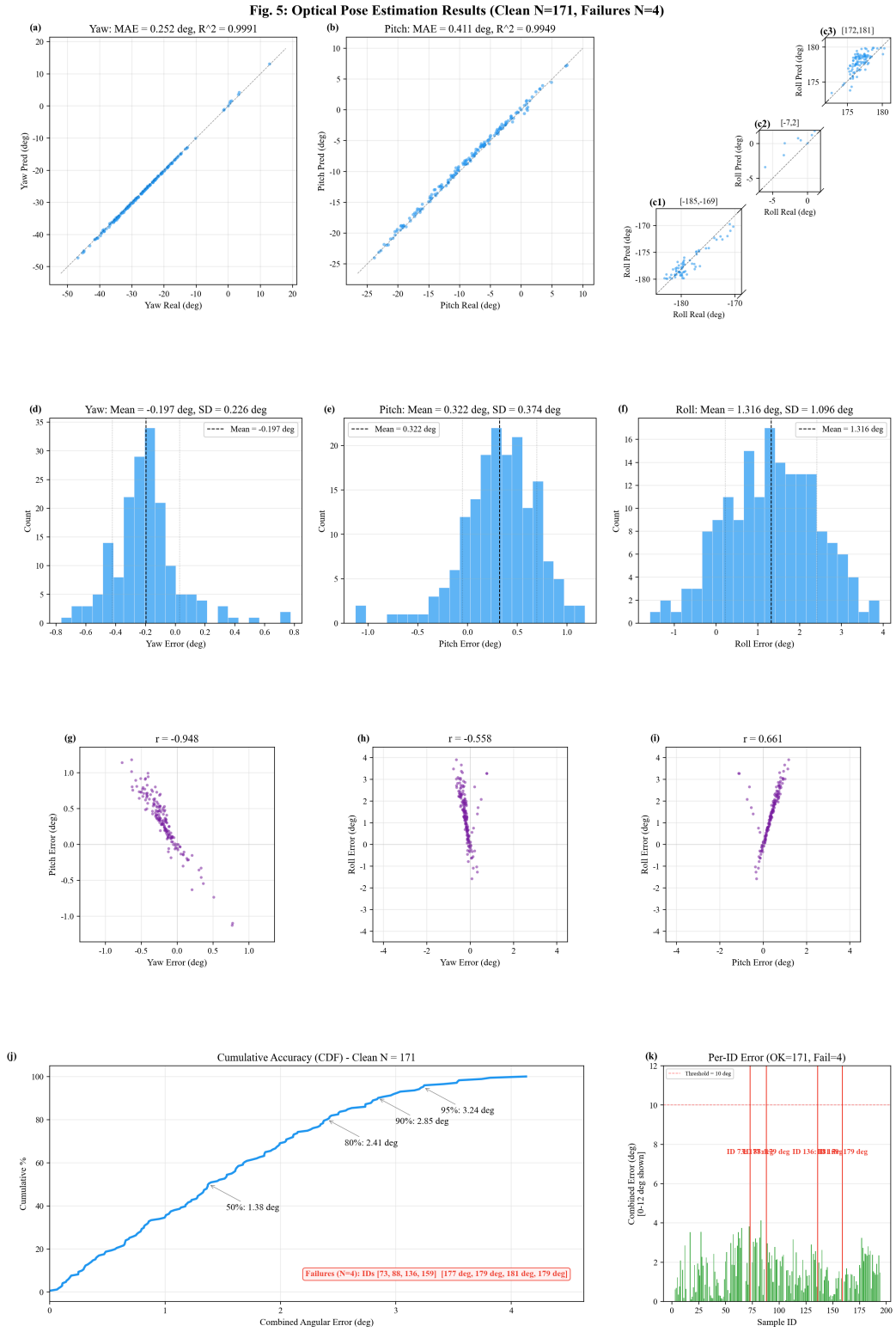


Figure 5

Optical pose estimation results.

Overview: Of 194 optical images, 19 (9.8%) excluded (detection failure or < 3 keypoints). Among 175 valid comparisons, 4 (2.3%, IDs = 73, 88, 136, 159) exhibited complete pose loss (combined error > 170 deg). The remaining 171 samples (97.7%) constitute the reliable set.

(a) Predicted vs. real Yaw angle.MAE = 0.252 deg, SD = 0.161 deg, median = 0.234 deg, max = 0.774 deg. The most accurate of the three axes.**(b) Predicted vs. real Pitch angle.**MAE = 0.411 deg, SD = 0.273 deg, median = 0.363 deg.**(c1-c3) Roll angle with broken axes.**Data split into three clusters: negative (c1), near-zero (c2), positive (c3), each shown as a zoomed panel. Three panels arranged along a diagonal (bottom-left to upper-right), connected by a single dashed y = x line. MAE = 1.438 deg, SD = 0.930 deg, median = 1.307 deg. Roll accuracy is lowest, consistent with weaker geometric constraint of cube corner features under in-plane rotation.**(d) Yaw error distribution.**Mean = 0.00 deg, SD = 0.30 deg.**(e) Pitch error distribution.**Mean = 0.01 deg, SD = 0.49 deg.**(f) Roll error distribution.**Mean = -0.01 deg, SD = 1.72 deg. All three distributions are approximately symmetric and zero-centered; residual errors dominated by random noise.**(g) Yaw vs. Pitch error scatter.**Strong negative correlation (r = -0.948). Overestimation of Yaw systematically accompanied by underestimation of Pitch, indicating a directional trade-off in the pitch-yaw plane estimate.**(h) Yaw vs. Roll error scatter.**Moderate negative correlation (r = -0.558).**(i) Pitch vs. Roll error scatter.**Moderate positive correlation (r = 0.661).**(j) CDF of combined angular error e_theta = sqrt(e_Y^2 + e_P^2 + e_R^2).**Reliable samples (N = 171): 50th = 1.38 deg, 80th = 2.41 deg, 90th = 2.85 deg, 95th = 3.24 deg. Median = 1.38 deg. Four failures (red box) at 177-181 deg attributed to complete pose loss under extreme viewing angles.**(k) Per-sample combined angular error (ID 1-194).**Y-axis cropped to 0-12 deg for detail. Green = reliable (e_theta <= 10 deg), red = failure (e_theta > 10 deg). Detection success rate = 90.2% (175/194).

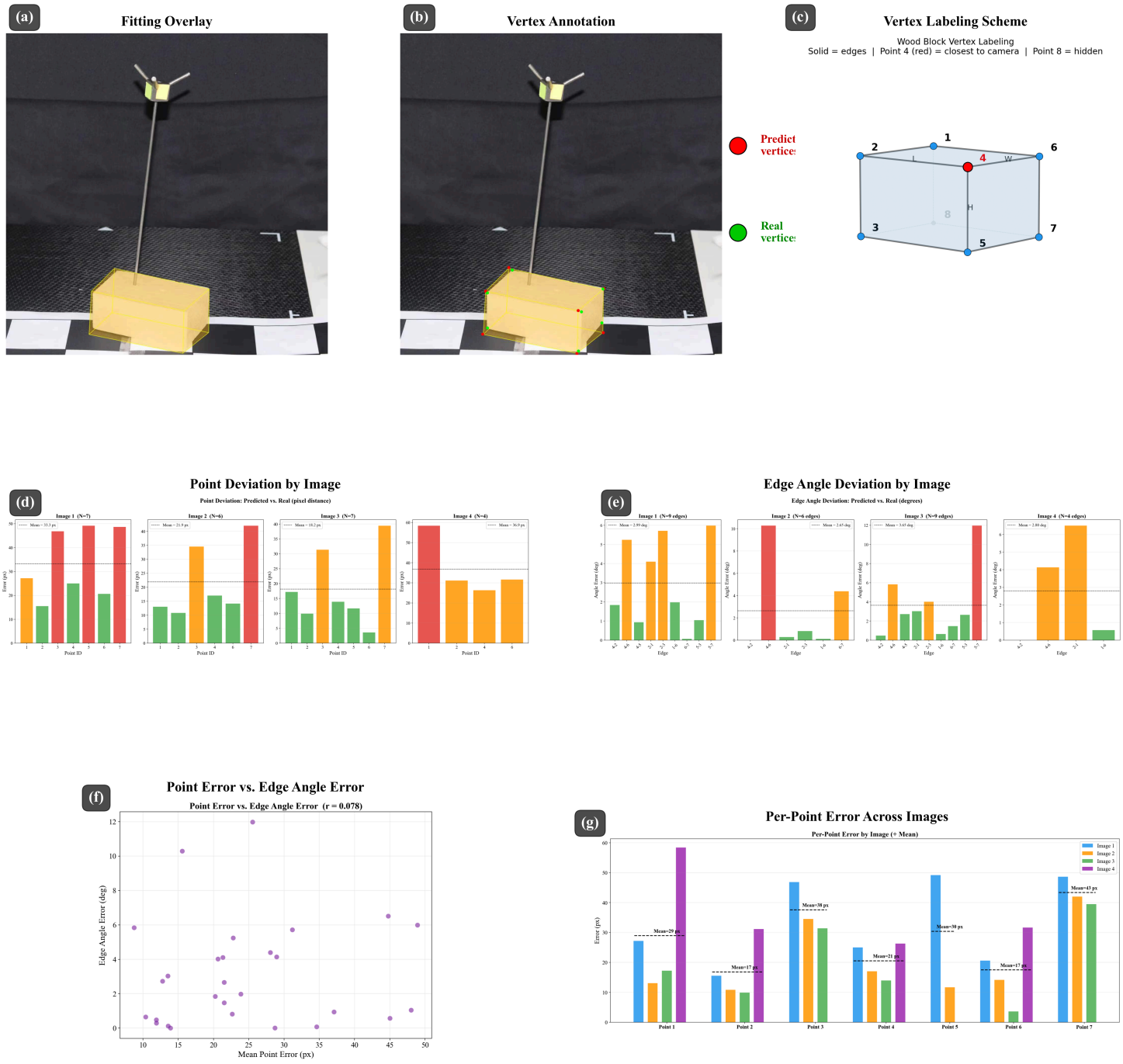


Figure 6

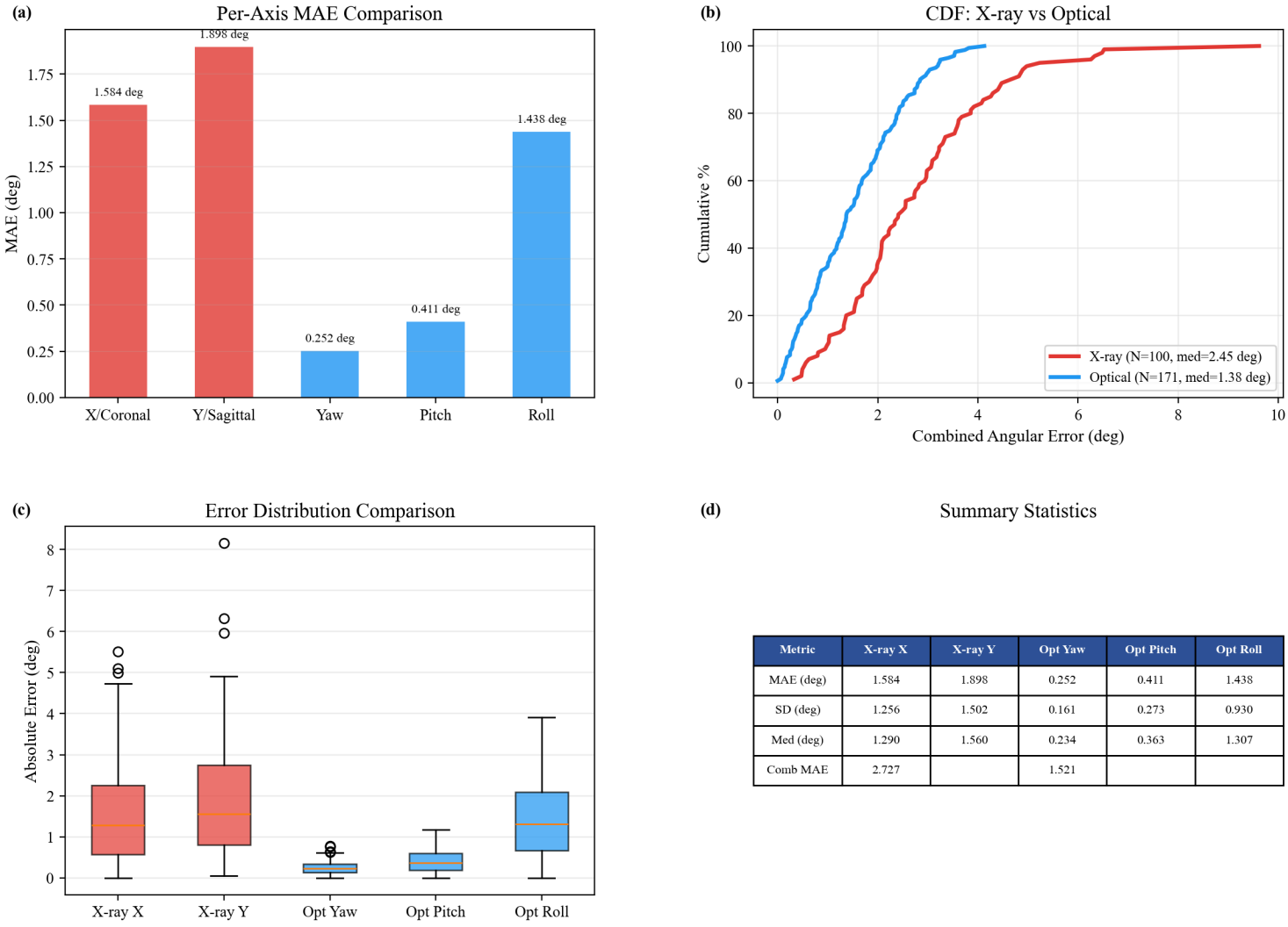
Proof-of-concept: pose transfer from marker to target object.

(a) Fitting overlay.A representative optical image showing the algorithm-predicted wood block wireframe overlaid on the real wood block. The marker (cube + four cylinders) is visible adjacent to the wood block. The wireframe is generated from the marker-to-wood-block relative pose estimated via the optical pose

pipeline.**(b) Vertex annotation.**The same image with manually annotated vertices. Red dots indicate predicted vertex positions derived from the algorithm output; green dots indicate the corresponding real wood block vertex positions identified by manual inspection. Seven of the eight vertices of the cuboid are visible (vertex 8 is occluded).

(c) Vertex labeling scheme. Three-dimensional diagram of the wood block showing the labeling convention for the eight vertices. Vertex 4 is the corner closest to the camera and serves as the reference point. The three edges emanating from vertex 4 define the length (L, to vertex 2), width (W, to vertex 6), and height (H, to vertex 5) of the block. Vertices 1, 3, and 7 lie on the face diagonals opposite vertex 4 on the three visible faces. Vertex 8 is the full 3D diagonal opposite vertex 4 and is occluded.**(d) Point deviation by image.**Bar chart of the Euclidean pixel distance between predicted and real vertex positions for each of the four images. Green = error < 25 px; orange = 25-40 px; red = > 40 px. Dashed line: per-image mean. Across all 24 point pairs (excluding missing annotations in Images 2 and 4), the mean point deviation was 26.6 px (SD = 14.9 px, median = 25.7 px, range = 3.6-58.4 px).**(e) Edge angle deviation by image.**Bar chart of the angular deviation (in degrees) between predicted and real edge direction vectors for each image. Only edges with both endpoints annotated are included (9 candidate edges per image; edges connected to vertex 8 are excluded). Green = < 4 deg; orange = 4-8 deg; red = > 8 deg. Across all 28 edge pairs, the mean edge angle deviation was 3.10 deg (SD = 3.09 deg, median = 2.32 deg, range = 0.00-11.97 deg).**(f) Point error vs. edge angle error.**Scatter plot of the mean point error of each edge (average of its two endpoint deviations) vs. the angular edge deviation. The near-zero correlation (r = 0.078) indicates that larger pixel deviations do not systematically translate into larger angular errors, suggesting that vertex prediction errors tend to manifest as approximately coherent translational shifts rather than independent random perturbations of individual vertices. **(g) Per-point error across images.**Grouped bar chart of point deviation by vertex ID, with individual bars for each of the four images and a dashed line indicating the cross-image mean. Vertices farther from the camera (1, 3, 5, 7) show systematically larger errors (mean = 29.0-43.4 px) than near-camera vertices (2, 4, 6; mean = 16.9-20.6 px). Point 7 exhibited the largest mean error (43.4 px, N = 3) and Point 2 the smallest (16.8 px, N = 4).

Fig. 7: Cross-Modal Comparison - X-ray vs Optical Pose Estimation



Metric	X-ray X	X-ray Y	Opt Yaw	Opt Pitch	Opt Roll
MAE (deg)	1.584	1.898	0.252	0.411	1.438
SD (deg)	1.256	1.502	0.161	0.273	0.930
Med (deg)	1.290	1.560	0.234	0.363	1.307
Comb MAE	2.727		1.521		

Figure 7

Cross-modal registration consistency.

(a) **Per-axis MAE comparison.**X-ray: X/coronal = 1.584 deg, Y/sagittal = 1.898 deg. Optical: Yaw = 0.252 deg, Pitch = 0.411 deg, Roll = 1.438 deg. Optical Yaw and Pitch substantially outperform X-ray axes (both < 0.5 deg), while Roll accuracy is comparable to X-ray Y-axis.**(b) Overlaid CDF of combined angular error.**X-ray (N = 100): median = 2.45 deg, MAE = 2.73 deg. Optical (N = 171): median = 1.38 deg, MAE = 1.52 deg. Optical modality achieves better overall accuracy with tighter distribution (SD: 0.98 deg vs. 1.58 deg).**(c) Boxplot of per-axis absolute errors.**Optical Yaw: most concentrated (IQR = 0.22 deg). X-ray Y-axis: largest dispersion (IQR = 2.19 deg). Optical Roll dispersion (IQR = 1.36 deg) notably smaller than both X-ray axes.**(d) Summary statistics for both modalities.**See table for full per-axis MAE, SD, median, and combined error values.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SI.docx