A metrological approach for Augmented Reality tooltip tracking assessment

Computers in Industry 175 (2026) 104430

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.sciencedirect.com/journal/computers-in-industry

A metrological approach for Augmented Reality tooltip tracking assessment

Federico Salerno *, Luca Ulrich, Giacomo Maculotti, Sandro Moos, Gianfranco Genta, Enrico Vezzetti, Maurizio Galetto

Department of Production and Management Engineering, Politecnico di Torino, Corso Duca degliAbruzzi, 24, Turin, 10129, Italy

ARTICLE INFO

Dataset link: link

Keywords:
Augmented reality
Tooltip tracking
Design of experiments
Measurement uncertainty
Metrological characteristics

ABSTRACT

Tracking systems are essential in various fields, such as health and manufacturing industries, enabling mapping between the real and digital worlds. Amongst others, Augmented Reality Tracking Systems (ARTS) are more recent and less explored. This work proposes a quantitative metrological methodology to evaluate ARTS tooltip tracking performance, facilitating benchmarking, parameter optimization, and system selection for specific tasks. A specific 3D-printed measuring artifact is proposed to guide tooltip positioning. Tracking accuracy and precision are estimated, highlighting the effects of influence factors. The methodology was tested with two commercial state-of-the-art ARTSs using marker-based tooltips, i.e., a Microsoft HoloLens 2 and a stereo camera system equipped with Intel RealSense SR305 cameras. Metrological characteristics are evaluated, and the Euclidean distance expanded uncertainty at a conventional 95% confidence level is estimated as 5.071 mm for the HoloLens 2 and 6.800 mm for the stereo system, resulting in a superior metrological performance of HoloLens 2 under the specified conditions. This study provides a standardized approach for quantitatively comparing AR tracking systems, offering valuable insights for optimizing their use in specific applications and, innovatively in the context of ARTS, associates measurement uncertainty with tracked distance values.

1. Introduction

Augmented Reality (AR) is a technology that improves real-world environments by overlaying digital information on the user's field of view (Feiner, 2002). This is typically achieved through devices such as smartphones, tablets, monitors, or head-mounted displays (Milgram et al., 1999), allowing users to perceive a seamlessly integrated blend of virtual and physical worlds (Milgram et al., 1995). AR systems enrich the user's perception by adding context-specific data (Rauschnabel et al., 2022), thereby expanding the interaction capabilities and sensory experience beyond the natural world (Scholz and Smith, 2016).

In the context of "augmented humans", AR plays a role alongside other enabling technologies such as advanced prosthetics (Innocente et al., 2022), neurotechnologies, and wearable devices (Pirmagomedov and Koucheryavy, 2021). AR's ability to provide real-time, context-sensitive information directly within the user's field of view can enhance cognitive and sensory functions (Dengel et al., 2021). For example, AR can assist in complex tasks by overlaying step-by-step instructions or critical data, thereby improving efficiency and reducing errors. Furthermore, when combined with other technologies such as brain-computer interfaces and smart wearables, AR can offer a better integrated augmentation experience, contributing to the enhancement of human capabilities (Raisamo et al., 2019).

The digitization process is central to AR functionality (Masood and Egger, 2020). This process converts analog features into digital formats by translating physical points into digital coordinates, capturing their 3D locations relative to a specific reference system. The digitization process is pivotal in AR to achieve accurate superimposition and spatial coherence (Yenduri et al., 2024) of virtual models, so that the AR applications can precisely augment the real world with virtual elements (Cannavò et al., 2020), enhancing both the consistency and the immersivity level of the technology. 3D acquisition exemplifies digitization, generating 3D point clouds from objects or scenes to create virtual models (Bi and Wang, 2010). Digitization process is crucial for reverse engineering, quality control, artwork reconstruction in the domain of cultural heritage (Innocente et al., 2024), and providing dynamic mapping for autonomous vehicles (Ulrich et al., 2021), as well as AR and Virtual Reality (VR) applications. In surgical procedures, AR enhances precision by overlaying digital anatomical structures onto patients (Birlo et al., 2022), aiding in the exact placement of surgical instruments like needles or drills (Vávra et al., 2017), reducing invasiveness and improving safety and recovery. In manufacturing, digitizing specific points integrates AR for non-destructive inspections (Amza et al., 2018), guiding assembly lines and enhancing quality control (Ho et al., 2022) with real-time, visually augmented

* Corresponding author.
E-mail address: federico.salerno@polito.it (F. Salerno).

https://doi.org/10.1016/j.compind.2025.104430
Received 7 September 2024; Received in revised form 25 September 2025; Accepted 24 October 2025
Available online 6 January 2026

feedback (Ramírez et al., 2015). Similarly, in robotics and automation (De Pace et al., 2018), specific holograms could be augmented, i.e. superimposed, in the environment (Zhu et al., 2019), improving navigation and safety, by supporting precise object positioning, avoiding collisions and streamlining workflow through visual integration of digital and physical realms.

Literature shows that adopting a tool and continuously tracking its tip is the widest adopted technique to retrieve the 3D coordinates of specific points in a scene with respect to a given reference system. There are two kind of tracking technologies (Franz et al., 2014): Electro Magnetic Tracking System (EMTS) and Optical Tracking System (OTS).

EMTS detects the 3D position of sensors within a magnetic field. Magnetic field generators can be AC-driven, using alternating current for a varying field with specific coil arrangements, DC-driven, with direct current for a steady field requiring different tracking methods (Yaniv et al., 2009), or passive, using permanent magnets for a constant but less flexible field. Electromagnetic tracking sensors, containing coils or magnetic materials, respond to field changes to determine their position. No line-of-sight is needed between sensors and generators, making them suitable for tracking internal instruments like endoscopes and needles (Chen et al., 2020).

On the other hand, OTSs use cameras to localize markers, requiring line-of-sight. There are two categories of optical tracking: pattern-based and localization-based. Pattern-based tracking uses AR fiducial markers. Among the most widely adopted are: ARToolKit, ARTag, AprilTag, ArUco, and QR Codes (Jurado-Rodriguez et al., 2023). Image processing identifies these tags and their corners, which are designed to be rotationally invariant for accurate tracking. Using the centers of at least 3 tags or known 3D points and their corresponding 2D image positions, tracking transformations are deduced using the Perspective-n-Point (PnP) algorithm (Nakano, 2016), allowing for effective tracking with a single camera. Localization-based tracking uses localization markers, distinct in color or reflectance, facilitating their centroid 3D localization, through thresholding and contour detection techniques (Parker, 2010). Markers can be active, thus illuminated by a light source (Korak and Kucera, 2015), or passive, typically retroreflective or colored. A specific pattern of localization markers is used with a stereoscopic camera setup (Yang et al., 2012) to perform the triangulation and compute the spatial positioning; subsequently, the tip position is determined through Iterative Closest Point (ICP) alignment using a mono camera and the PnP algorithm. If at least three markers are rigidly attached to a tool, its tip can be 3D tracked. Despite the efforts to control partial occlusions (Garrido-Jurado et al., 2014; Wang et al., 2015, 2023), optical tracking systems require a line-of-sight, posing challenges in crowded spaces like operating rooms and precluding their use for tracking flexible or small instruments. During the tracking process, the desired tooltip 3D coordinates are collected and the feedback can be visualized (Zollmann et al., 2020) through a point cloud display.

While Augmented Reality Tracking Systems (ARTSs) can be considered a subset of Optical Tracking Systems (OTSs), as tracking is always performed using cameras (Ulrich et al., 2024), real-time visualization is a distinguishing feature of ARTSs. This is achieved by superimposing the tooltip position onto the physical object using a monitor (Alves et al., 2022) or a Head Mounted Display (HMD) device.

Correct augmentation of the virtual content are core to develop a solution capable of satisfying the constraints required into applications such as surgery (Radkowski and Kanunganti, 2018; Stradiotti et al., 2024) and manufacturing (Settimi et al., 2022; Eck et al., 2015).

This work aims to provide a metrological characterization of tracking performances through metrological characteristics and uncertainty evaluation of ARTSs. Using a Design of Experiments (DOE) approach, we identify significant factors and introduce a cost-effective measuring artifact. This artifact features various geometric shapes designed to thoroughly test tracking accuracy under different conditions, offering a practical tool for assessing precision and accuracy in applications involving ARTS technologies. In particular, the key contributions of this study are outlined as follows:

1. Metrological performances evaluation: this study presents a comprehensive evaluation protocol based on a well-established metrological approach to assess tooltip tracking in digitization processes for ARTS. The approach involves a detailed analysis of measurement uncertainty contributions to benchmark the tracking methods and hardware performance, to evaluate their precision and accuracy against established measurement standards.
2. Design of Experiments (DOE) approach to identify statistically significant tracking influence factors: to identify key influence factors through the metrological methodology, the study utilizes a DOE approach. This enables systematic investigation into how different factors impact the accuracy and efficiency of tooltip tracking, thereby enhancing the reliability of the findings.
3. Measuring artifact: the study proposes an alternative artifact specifically designed to assess the impact of geometric factors on tooltip tracking. Suitable for fabrication using standard desktop 3D printers, it incorporates various geometric shapes to challenge the tracking system in diverse locations and orientations of the tracked tool.

The effectiveness of this approach is demonstrated through evaluations using two ARTS: the AR HMD Microsoft® HoloLens 2 and a custom-built monitor-based AR device, using a stereoscopic system with Intel RealSense SR305 RGB cameras, targeting different marker-based tooltips to verify precision in various applications. As will be further clarified throughout the paper, to validate this framework, we conducted evaluations using two distinct tracking systems, differentiated both by their hardware configurations and underlying software algorithms. The objective was to demonstrate the adaptability and generalizability of our approach, which allows for a comparison of heterogeneous tracking technologies.

The paper is organized as follows: Section 2 reviews related works on tooltip tracking and measurement artifacts. This section highlights both the focus on tracking performance evaluation and the general lack of a metrological framework for the evaluation approach. Section 3 presents the methodology applied to ARTS and includes two case studies to test it. Furthermore, Section 4 reports the results of the methodology application. Section 5 provides a critical analysis of the results and a comparison with the values reported in the literature, while Section 6 draws the conclusions. The complete results are presented in Appendix A.

2. Related works

Various studies in the literature have adopted measuring artifacts to quantitatively assess localization precision and accuracy. In the context of EMTSs one of the first works, proposed by Day et al. (1998) in 1998, evaluated the accuracy of the Optotrak system (Northern Digital Inc., Waterloo, Ontario, Canada) using a testing artifact consisting of a Plexiglass panel with adjustable holes on a slotted wooden base (Fig. 1a). In the same year, Birkfellner et al. (1998) assessed the accuracy of an AC-driven Isotrak II (Polhemus Inc., Colchester VT, USA) and a DC pulsed system (The Bird, Ascension Inc., Burlington VT, USA), employing a Plexiglass Cartesian positioner as a testing artifact (Fig. 1b). In 2003 Frantz et al. (2003) used a translation table equipped with a sensor capable of rotating around its axis to evaluate the NDI Aurora system, an electromagnetic tracking system used in medical applications consisting of a System Control Unit, Field Generator, Sensor Interface Unit, and sensors (Northern Digital Inc, 2020) (Fig. 1c). These components enable real-time tracking of instruments without requiring a direct line of sight, facilitating precise localization and visualization of tools during procedures such as interventional radiology, electrophysiology, and endoscopy. Some years later, Hummel et al. (2005) developed a renowned testing artifact comprising a polycarbonate panel with machined holes to test the accuracy of

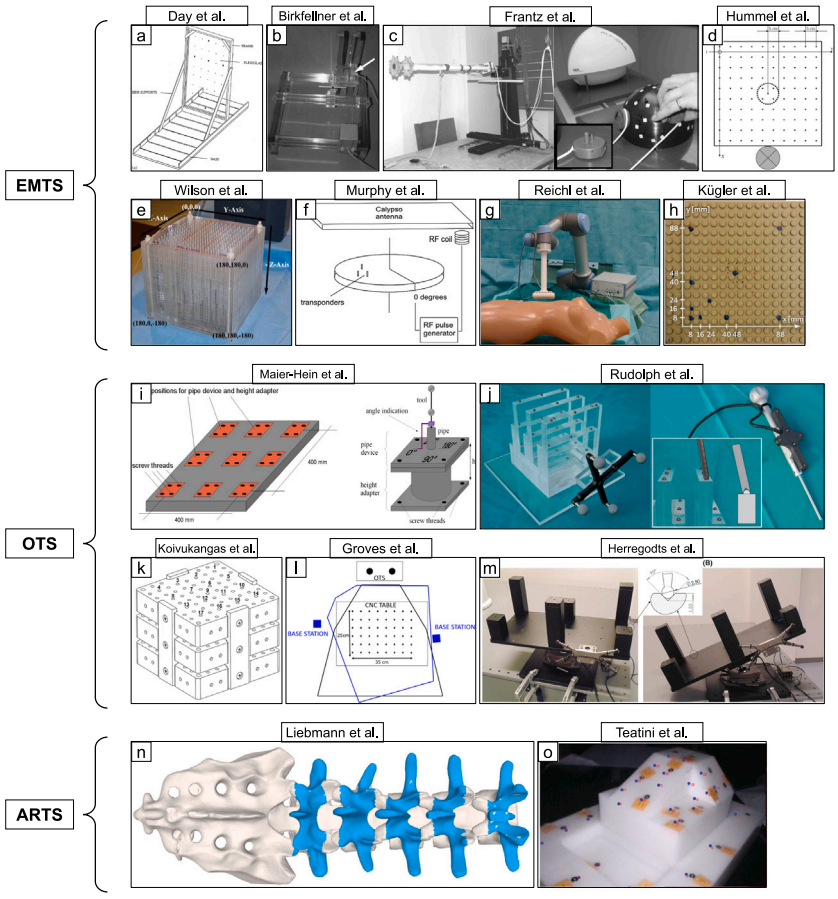

Fig. 1. Related works artifacts.

NDI Aurora and Ascension microBIRD (Fig. 1d). Analogously, Wilson et al. (2007) introduced a Plexiglass cubic artifact with holes of varying depths to assess the NDI Aurora, instead of positioning the Hummel plate at various heights (Fig. 1e). In 2007, Shen et al. (2007) utilized a mechanical three-axis positioning robot to assess the tooltip tracking accuracy of the NDI Aurora system.

Similarly, Atuegwu and Galloway (2008) employed a Plexiglass panel artifact with machined holes to evaluate the NDI Aurora, akin to the method used by Hummel et al. Murphy et al. (2008), who evaluated the Calypso system (Calypso Medical Technologies, Seattle, WA) using a rotating table equipped with transponders (Fig. 1f). Santanam et al. (2009) assessed the Calypso system using a test box, which was moved to different points on a fixed table, while Haidegger et al. (2011) employed a LEGO platform and a sensor mounted on a brick tower, moving it to different positions on the platform, to evaluate three EMTS: Ascension's 3D Guidance medSAFE (DC field), and both 5 DOF and 6 DOF sensor NDI Aurora. Gergel et al. (2012) developed a tripod robotic arm, equipped with a sensor on its hand effector, to locate the sensor in different positions within the workspace and investigate two configurations of NDI Aurora: planar and tabletop field generator. In 2013 Reichl et al. (2013) tested NDI Aurora's accuracy by moving a sensor with a six-axis robotic arm (Fig. 1g) and in the same year Yoo et al. (2013) compared the NDI Aurora field generator with a configuration where the generator was embedded in the operating table, employing a Cartesian system to maneuver sensors within the working volume. Some time later Lugez et al. (2015) assessed the NDI Aurora V2 (AC field) using 3D printed cubic and hemispherical artifacts to evaluate position and orientation accuracy, respectively. Kügler et al. (2019) tested the NDI system using a LEGO platform and a sensor mounted on a brick tower, moving it to various positions on the platform (Fig. 1h). In 2021, Lee et al. (2021) evaluated the accuracy of the IGESS Medigator® system, an image-guided endoscopic sinus surgery system that integrates the NDI Aurora for navigation during sinus surgeries. Tracking accuracy was assessed using a 3D-printed phantom designed according to the ASTM F2554-10 standard (ASTM International, 2010). The phantom contained cone-shaped divots as fiducial markers, and a tracked pointing tool was used to measure the positions of these divots.

Moving to OTSs, in 2008 Maier-Hein et al. (2008) assessed the NDI Polaris system with passive optical localization markers and the MicronTracker 2 H40 (Claron Technology, Inc., Toronto, Ontario, Canada) with pattern-based markers, using a measuring artifact consisting of a base panel where a support could be mounted with screws, allowing the stylus to rotate around its axis (Fig. 1i). In 2010 Rudolph et al. (2010) evaluated the accuracy of three OTSs: two Optotrak 3020 s, a Polaris P4, and a Polaris Spectra system, with both Polaris systems using passive optical localization markers, employing an acrylic glass artifact with locating spheres (Fig. 1j). Some years later, Koivukangas et al. (2013) used a multilevel artifact with machined holes to assess the accuracy of interchangeable use of OTS and EMTS navigators, specifically the StealthStation S7 (Medtronic Inc., Louisville, CO, USA) (Fig. 1k). More recently, Groves et al. (2019) utilized a CNC table and a hemispherical artifact with holes to evaluate the accuracy of NDI Polaris, HTC VIVE, and a hybrid tracking system developed by the authors (Fig. 1l). In 2020 Herregodts et al. (2021) investigated the accuracy of three OTSs: Krypton K610 (Metris), Vicon Vantage V16 (Vicon), and OptiTrack Flex 13 (NaturalPoint), with the first and second

Table 1
Tooltip tracking accuracy assessment in the literature. EMTS: Electromagnetic Tracking System; OTS: Optical Tracking System; ARTS: Augmented Reality Tracking System. The "Qualified" column specifies whether the measurement artifact was preliminarily qualified using a measurement instrument.

	Author	Year	Measured artifact	Qualified
EMTS	Day et al. (1998)	1998	Adjustable drilled plexiglass panel to pre-set positions	Yes
	Birkfellner et al. (1998)	1998	Plexiglass Cartesian system	Yes
	Frantz et al. (2003)	2003	Rotating cylinder on a translation table	Yes
	Hummel et al. (2005)	2005	Polycarbonate panel with machined holes	Yes
	Wilson et al. (2007)	2007	Plexiglass cube with machined holes	Yes
	Shen et al. (2007)	2007	Mechanical three-axis positioning robot	No
	Atuegwu and Galloway (2008)	2008	Plexiglass panel with machined holes	No
	Murphy et al. (2008)	2008	Turntable with transponders	No
	Santanam et al. (2009)	2009	Localization test box	No
	Haidegger et al. (2011)	2011	LEGO platform with sensor mounted on a brick tower	No
	Gergel et al. (2012)	2012	Tripod calibration robot	No
	Reichl et al. (2013)	2013	Six-axis robotic arm	No
	Yoo et al. (2013)	2013	Wooden Cartesian system	No
	Lugez et al. (2015)	2015	3D printed cube and emispherical artifact with holes	No
	Kügler et al. (2019)	2019	LEGO platform with sensor mounted on a brick tower	No
	Lee et al. (2021)	2021	ASTM F2554-10 phantom (ASTM International, 2010)	Yes
OTS	Maier-Hein et al. (2008)	2008	Base panel holding a rotating pipe device	No
	Rudolph et al. (2010)	2010	Acrylic glass artifact with locating spheres	No
	Koivukangas et al. (2013)	2014	Multilevel artifact with machined holes	Yes
	Groves et al. (2019)	2019	CNC table and emispherical artifact with holes	Yes
	Herregodts et al. (2021)	2020	Tiltable base with different hight machined divot	No
ARTS	Liebmann et al. (2019)	2019	Anatomical 3D printed phantom	No
	Kunz et al. (2020)	2020	Robotic arm	No
	Teatini et al. (2021)	2020	Phantom with horizontal and tilted surfaces	No
	Martin-Gomez et al. (2023)	2022	3-axis linear translation stage + rotation platform	No

systems based on passive IR retroreflective markers, and the third on active IR markers; their measuring artifact included a metal plate with various metal towers fixed to it and drilled cones divots scattered on both the surface of the metal plate and the tops of the towers (Fig. 1m).

Concerning ARTSs, in 2019 Liebmann et al. (2019) utilized the Microsoft® HoloLens 1, employing it as an inside-out technique. This approach uses the HMD both as a tracker and a display system. A pointing tool equipped with an AprilTag marker was tracked during a digitization procedure applied to anatomical phantom artifacts (Fig. 1n). Furthermore, in 2020 Kunz et al. (2020) evaluated the tracking accuracy of single IR retroreflective passive markers. They used a combination of "short throw reflectivity" and "short throw depth" video sources, as well as a stereo rig with left front and right front cameras. A robotic Stewart platform (Hexapod, PI, Germany) was used to move the markers. In the same year, Teatini et al. (2021) studied the accuracy of a hand-eye AR system, consisting of a tracked stereo laparoscope ENDOEYE Flex3D (Olympus, Tokyo, Japan), and the NDI Polaris Spectra (Waterloo, Canada). Hand-eye calibration between the laparoscope and the tracked tools was achieved using passive localization markers; a polyoxymethylene artifact with horizontal and tilted surfaces with machined divots was used for accuracy assessment (Fig. 1o). In 2023, Martin-Gomez et al. (2023) employed the combination of reflectivity and depth stream. To precisely control the movements of the markers, a 3-axis linear translation stage and a rotation platform were used. Building on these contributions, more recent studies have further explored the use of AR-based tracking systems in different contexts. For instance, Narasimhan et al. (2025) in 2025 developed a multi-stage pipeline for monocular 3D tooltip tracking in robotic surgery, combining zero-shot approaches with supervised learning to evaluate precision and robustness in realistic scenarios. In the same year, Singh et al. (2025) investigated the real-world tracking accuracy and depth drift of XR devices, providing quantitative insights into positioning errors and stability that are directly relevant to AR tool-tracking applications. Similarly, Kavitha et al. (2025) presented a comparative study of surgical tool tracking using AR, where a stereo camera marker-based setup was evaluated against an infrared optical tracking system (OptiTrack), highlighting differences in accuracy and precision.

Tool-tracking is a well-known and established procedure, as evidenced by the literature analysis. Among the three main tracking solutions, EMTSs can be considered a mature and reliable technology. Studies evaluating their accuracy date back to the late 1990s (1998), and interest in them remains high. In contrast, the main accuracy evaluation studies for OTSs, using custom-made measuring artifacts, began about 10 years later (2008) and are less common. Furthermore, due to the recent emergence of AR technologies and the complexity of the required hardware and software, there are fewer and more recent studies (starting from 2019) on the accuracy evaluation of tooltip tracking using ARTS (Fig. 2). Moreover, it seems that a metrological characterization of the tooltip tracking error applied to ARTS has not yet been addressed.

Since the measuring artifact serves as the ground truth, it is recommended to perform a preliminary metrological qualification using tools like a Coordinate Measuring Machine (CMM) or a 3D scanner. Relying solely on the declared fabrication tolerance of the manufacturing machine may not be sufficient for the metrological qualification of the artifact. Table 1 summarizes these details: the "Qualified" column indicates whether the measurement artifact has undergone a qualification procedure. Although many studies have proposed methods for testing tooltip tracking systems, there is often a lack of a metrological approach that addresses the qualification of measuring artifacts, details their metrological characteristics, identifies key factors, and estimates uncertainty propagating relevant contributions. Additionally, while many works have introduced specific calibration artifacts for EMTS, less focus on OTS and limited attention to Augmented Reality Tracking Systems (ARTS) technologies was paid.

3. Materials and methods

A preliminary assessment procedure helps to evaluate the system's performance under specific operational conditions (Wiles et al., 2004). This evaluation includes the preliminary selection of working variables such as the employed hardware, illumination, and the distance between the tool and the device and the identification of the typical shape and

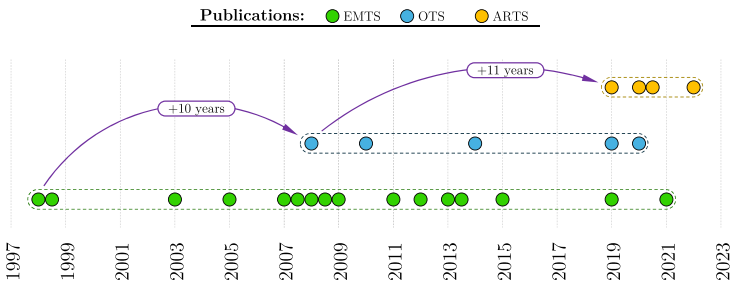

Fig. 2. Temporal distribution of related works.

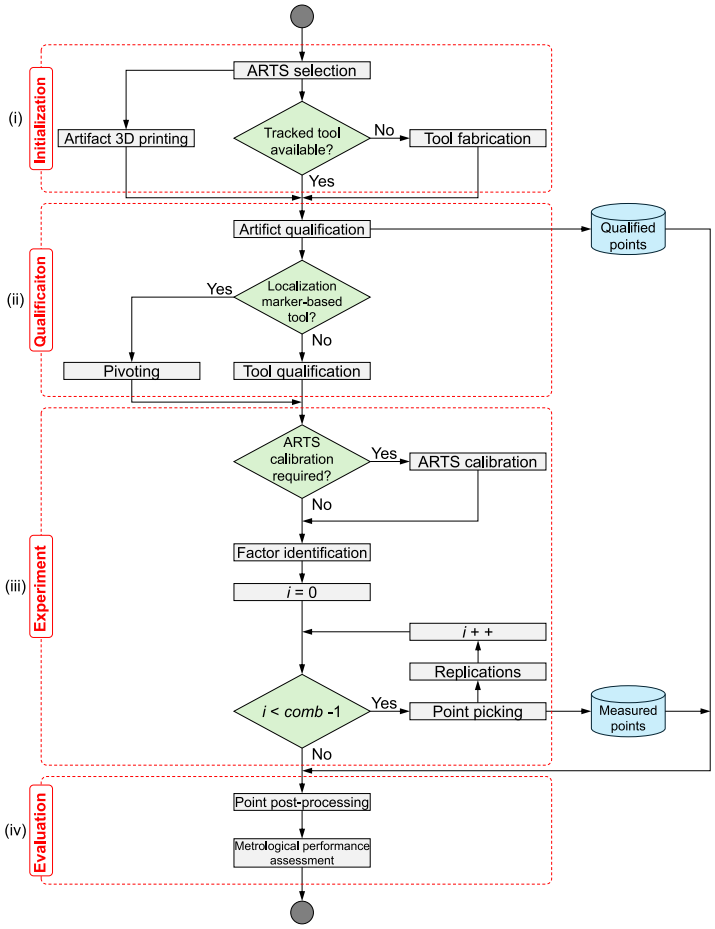

Fig. 3. Proposed workflow for the ARTS tooltip tracking metrological assessment, which includes four stages: initialization, qualification, experimentation, and evaluation.

dimensions of the measured objects, i.e., the measurand. Subsequently, understanding the statistically significant factors can provide guidance in optimizing the tracking system itself.

The metrological procedure allowing for the evaluation of both significant influence factors and metrological performances is schematized in Fig. 3. The procedure consists of four stages: (i) the initialization, which includes the ARTS selection, the fabrication of the artifact and tracking tool, (ii) the qualification of both artifact and tool, (iii) the experiment with the point-picking procedure, and (iv) the evaluation of the performances using collected measures. These stages are described in the following four subsections, while the fifth subsection concerns the case study, i.e. application of the proposed protocol.

3.1. Initialization

The first stage consists of the fabrication of the measuring artifact and the selection or, eventually, the creation of the tracked tool. Although the literature has proposed several artifacts, such as those shown in Table 1, for the evaluation of point-picking accuracy, we propose a specific one with both simple and various geometrical features. The design combines features of alternatives proposed in the literature, while being inspired from benchmarks used to test the performance tolerances of Additive Manufacturing (AM) machine (Minetola et al., 2020), as shown in Fig. 4a With a square base, measuring 200 mm on each side, the artifact is specifically designed for Material Extrusion AM process (ISO/ASTM, 2021) because its walls have angles that do

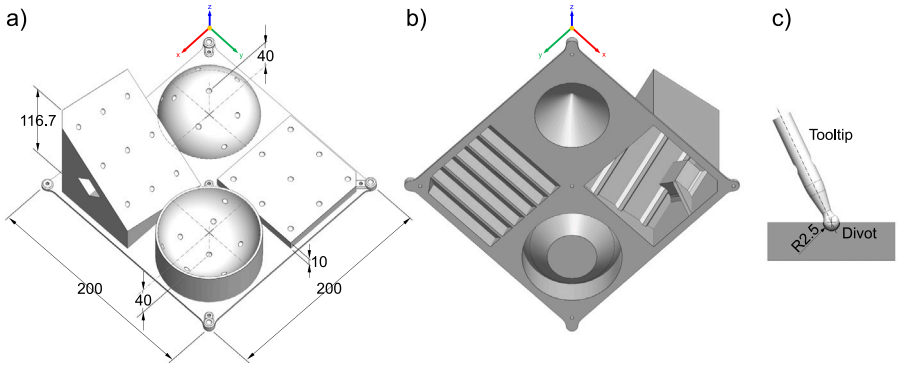

Fig. 4. Proposed artifact for tooltip tracking accuracy assessment. (a) Isometric view; (b) Bottom view; (c) Detail of tooltip and divot.

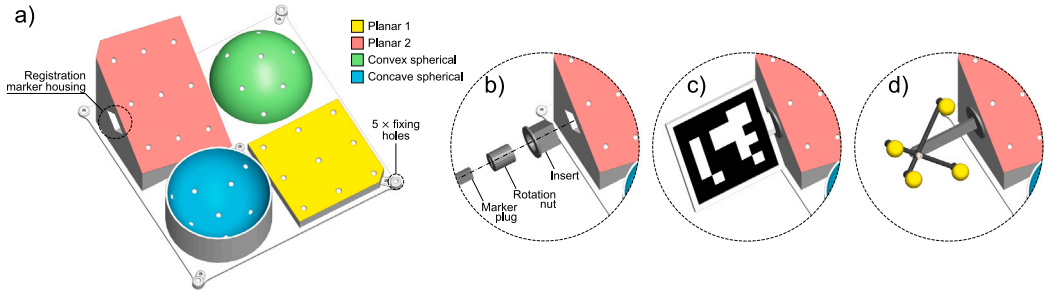

Fig. 5. (a) Characteristic geometries of the artifact; (b) Slot and components to accommodate an auxiliary reference system; (c) Pattern-based reference system; (d) Localization marker-based reference system.

not exceed 45° from the vertical plane (Fig. 4b). The artifact features 36 divots which are hemispherical concave surfaces with a radius of 2.5 mm. Accordingly, a tool with a spherical tip of 2.5 mm radius could fit into these divots and pivot.

The artifact features four different surfaces with divots: a planar horizontal surface, a planar tilted surface, a hemispherical convex surface, and a hemispherical concave surface. The design, depicted in Fig. 5a, has been selected to mimic various working conditions, additional to a simple, flat planar surface. A designated slot within the artifact is provided to house an auxiliary reference system, to which the points captured during the tracking test can be aligned. Fig. 5b, c, and d illustrate the mechanical integration of the auxiliary reference system, including pattern-based and localization marker-based systems, respectively.

This stage provides the foundation for the subsequent qualification, ensuring that the designed components are ready for metrological assessment.

3.2. Qualification

The qualification process requires prior camera calibration to determine the intrinsic and extrinsic parameters of the camera (Zhang, 2000) if these parameters are not available from factory data or the datasheet. This calibration can be accomplished using patterns such as checkerboards, Charuco boards, or circle grids (MathWorks, 2024a; OpenCV Community, 2024). An ARTS relies on processing video streams from cameras using these intrinsic and extrinsic parameters.

The process also involves capturing the actual geometries of both the artifact and the tracking tool. Due to variations in the manufacturing process, the 3D positions of the divots on the artifact may differ from those in the CAD model. Consequently, the qualification process aims to determine the actual coordinates of the divot centers in the reference system of the measuring instrument used for the artifact calibration. This measuring instrument could be any Coordinate Measuring Systems (CMS) (ISO, 2020), either contact-based, such as a Coordinate Measuring Machine (CMM), or non-contact, like a laser scanner. Regarding the tracked tool, the tracking process always involves superimposing a virtual model of the tool onto the real one, regardless of the presence and nature of markers. Correct tooltip tracking is achieved by superimposing the real geometry of the tool due to the geometrical distortions caused by the fabrication process. If it is possible to measure the 3D geometry of the tracking tool, the real position of the marker rig and the relative tip position can be retrieved; otherwise, the pivoting calibration procedure (Yaniv, 2015) can be used. The pivoting calibration is suitable for marker-based ARTS and consists in pivoting the tooltip into a divot and rotating the tool around the tip so that the marker rig lies on a trajectory sphere concentric to the divot itself. In the following, the notation used on the University of Illinois website on Robotic Systems (Illinois Motion Group, 2024) will be adopted: given two reference frames $\{A\}$ and $\{B\}$, the coordinate transformation from $\{A\}$ to $\{B\}$ is expressed as $\mathbf{T}_A^B$. The notation $\mathbf{p}^A$ indicates that the generic point $\mathbf{p}$ is expressed with respect to the reference frame $\{A\}$. During the tool pivoting procedure, the transformation matrix $\mathbf{T}_M^D$ of the marker rig reference system M, with respect to the ARTS reference system $\{D\}$, i.e., the pose of the marker rig relative to $\{D\}$, is estimated for each sampling instance. Since each marker describes a sphere and all spheres are concentric, the common center of the spheres $\mathbf{c}^D$, with respect to $\{D\}$, is calculated. Furthermore, it is possible to calculate $\mathbf{c}^M$ using the following relationship:

$$\mathbf{c}^D = \mathbf{T}_M^D \cdot \mathbf{c}^M \tag{1}$$

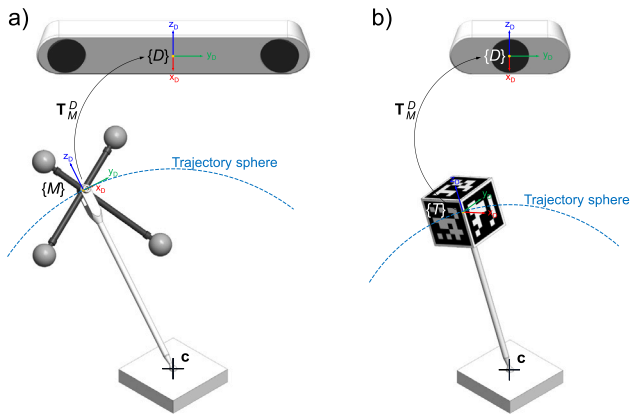

Fig. 6. Pivoting calibration procedure of two different tooltips. (a) Localization marker-based; (b) Pattern marker-based.

Here, $\mathbf{c}^M$ is the tooltip position relative to $\{M\}$, known as the "tooltip offset" (Fig. 6).

In our study, the optical 3D CMS (Geometrical, 2021) ATOS ScanBox 4105 (ZEISS® INDUSTRIAL QUALITY SOLUTIONS, Oberkochen, Germany) was used for the qualification process, in conjunction with GOM Inspect Pro software. Each hemispherical divot was modeled using a Gaussian sphere. With the artifact and tool qualified, the experiment can be carried out on reliable geometrical references.

3.3. Experiment

A Design of Experiments (DOE) is defined to allow for the identification of statistical significance of factors on tracking position. This is expedient in assessing systematic and random errors to allow uncertainty evaluation of the tracking, thus enabling performances comparison of different tracking systems. Thus, it is necessary to identify the influence factors of the phenomenon. This identification is required both for a comprehensive understanding of the phenomenon and to find and optimize the results. According to the authors' experience and literature analysis, there are five main families of influence factors, each containing specific factors related to the tooltip tracking process. An Ishikawa diagram shows main influence factors in Fig. 7.

AR interaction dynamics make up the first family, encompassing the key features of how users interact with the AR application. This includes the feedback mechanisms, such as sounds or video events triggered by user actions, and how augmented content is displayed, whether as a transparent mesh, wireframe, or superimposed on the real target, partially or fully occluding it. The type of gesture required to trigger an event is also significant, as it directly relates to the feedback mechanism. Ergonomics is part of this family as well: AR devices can be wearable, like HMDs, or handheld, such as smartphones and tablets. They might integrate the tracking system (inside-out tracking) or use separate devices for tracking and display (outside-in tracking). The distance between the tracked tool and the AR tracking device could influence the tracking process itself as well.

The second family pertains to the tracking system setup, which includes specific tracking system solutions such as camera configuration, e.g. monocular vision or stereoscopy, and camera specifications, primarily resolution (Sorriento et al., 2019). This family also includes the tracking algorithm, whether computer vision or machine learning methods are used, and the detection principle related to the specific nature and arrangement of markers. Specifically, this encompasses markerless tracking systems and the shape of the tool, considering both the spatial arrangement of the markers and the distance from the marker rig to the tip (Maurer et al., 1997).

The point-picking procedure is another source of variability. Its main identifiable factors are the orientation of the tool relative to the tracking system, the operator effect, tool holding, i.e., whether the tool is held in a fixed position or by hand, thus constantly moving, and the pose of the artifact with respect to the tracking system.

The fourth family pertains to the measurand: the type and number of geometric features, if any, the dimensions of the artifact, the shape of the divots (West et al., 2003; West and Maurer, 2004), and the accessibility of specific features to measure.

On the other hand, the environment is a well-known, less controllable aspect, which includes illumination conditions, free line of sight between the tracking system and the tracked object, and the background, which may lead to poor detection or false detection of markers. For this reason infrared optical tracking systems are usually adopted, especially in the surgical field, to minimize the influence of visible light conditions.

Due to the complexity of the phenomenon, together with the need to identify the significance of the factors, the DOE approach is suitable for this analysis. The selected factors are: geometry, i.e., the geometrical features of the measured artifact, detailed in Section 3.1, rotation, i.e. the orientation of the artifact with respect to the OTS, tool-device distance, and the operator, i.e. the specific participant involved in the digitizing procedure.

In this work, on one hand, for the sake of simplicity and to contain experimental effort, more influence factors could have been included in the measurement. On the other hand, while many aspects have been considered separately for clarity, they are inherently linked. For instance, each operator holds the tool differently, with varying levels of stability, and orients it uniquely relative to the tracking system. Furthermore, the tool holding and orientation can be affected by the specific AR interaction dynamics. The relative orientation and distance between the artifact and the operator, including the pose and accessibility, ultimately depend on the specific geometry of the measurand feature to be tracked. More specifically, accessibility is related to geometry, as a specific shape may be more or less reachable with the tool. Additionally, changing the pose can make certain parts of the measurand more accessible compared to a different pose, and vice versa. This work does not include any environmental factor in the DOE, limiting the scope to indoor applications (Huang et al., 2020) focusing on the contexts where these technologies are most commonly used (Huey et al., 2011). In medical settings, procedures are typically conducted under controlled, consistent conditions to ensure accuracy and safety (Smith and Schwiegerling, 2023). Similarly, manufacturing and assembly processes are generally carried out indoors to maintain precision and control over the working environment (Chandel et al., 2022). For this reason, factors such as lighting, occlusions, or reflections were excluded from the DOE, since the work focuses on controlled indoor conditions typical of AR tracking applications.

This results in a 4-factor design, with 2 qualitative factors, i.e., the geometry and the operator, and 2 quantitative, i.e., the rotation and the tool-device distance. Per each factor the levels shown in Fig. 8a are considered, leading to a full factorial design with $4 \times 2 \times 2 \times 3 = 48$ treatment combinations. For feasibility reasons, the operator factor was blocked, and 5 replications per condition were considered.

Each experiment replication involves positioning the tool in the divots by centering the spherical tip within the divot, ensuring that the convex surface of the tip and the concave seat are concentric. For each placement, the 3D coordinates of the tracked tip are collected and recorded. Each replication consists of initially collecting three points, L_1, L_2, and L_3, as detailed in (Fig. 8b) to establish a local reference system $\{L\}$. This is followed by the collection of 36 points, i.e., 9 for each geometry, totaling 39 points in one replication, thereby covering the four levels of the Geometry factor. Here, it is worth remarking that the 36 measured points on the geometrical features will be considered a nested factor in the geometry factor in the DOE analysis.

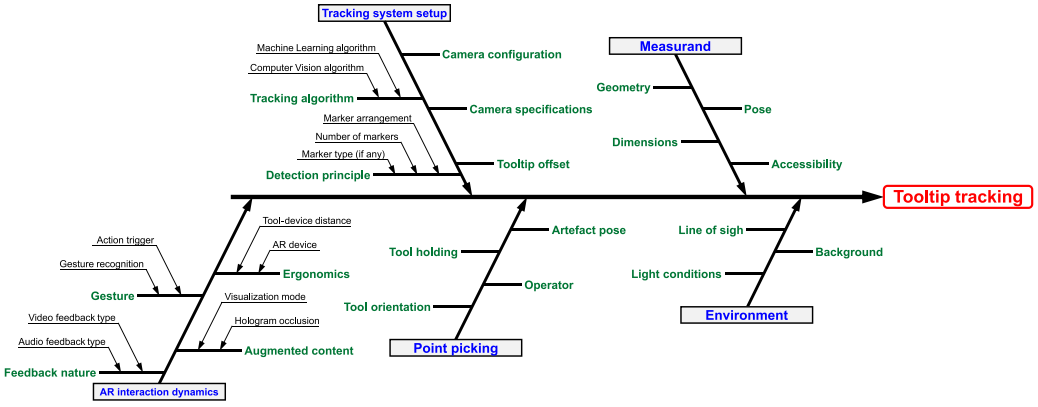

Fig. 7. Diagram of the influence factors of tooltip tracking process.

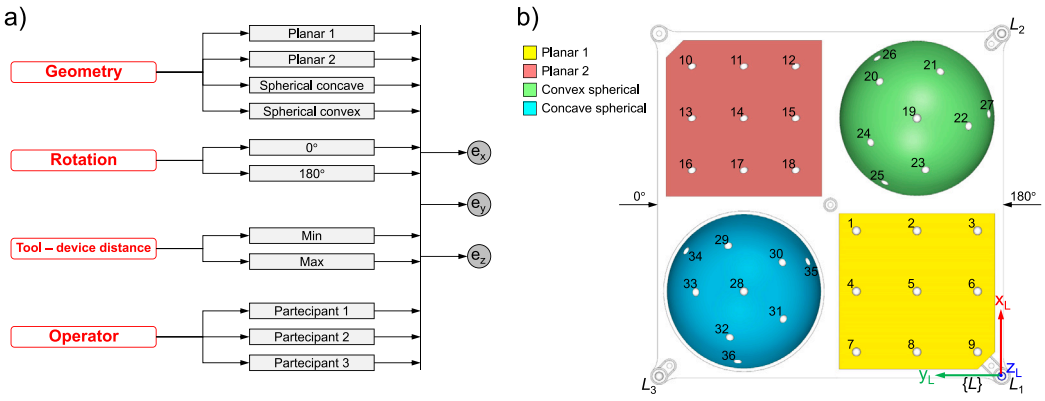

Fig. 8. (a) DOE overview: four factors, i.e., geometry, rotation, tool-device distance and operator with their levels produces a 48 treatment design with e_x, e_y and e_z response variables; (b) Point picking sequence and rotation: L_1, L_2, and L_3 determine a local reference system to which points from 1 to 36 are referred; arrows at 0° and 180° determine the orientation of the artifact with respect to the ARTS.

The local reference system $\{L\}$ is defined as follows: L_1 is the origin, the $\mathbf{x}_L$ axis is defined by the vector $\overrightarrow{L_1L_2}$, the $\mathbf{y}_L$ axis lies in the plane passing through L_1, L_2, and L_3, and is perpendicular to $\overrightarrow{L_1L_2}$ pointing towards L_3, while the $\mathbf{z}_L$ axis follows the right-hand rule:

$$\begin{cases} \mathbf{x}_L = \dfrac{\overrightarrow{L_1L_2}}{\|\overrightarrow{L_1L_2}\|} \\ \mathbf{y}_L = \mathbf{x}_L \times \mathbf{z}_L \\ \mathbf{z}_L = \dfrac{\overrightarrow{L_1L_2} \times \overrightarrow{L_1L_3}}{\|\overrightarrow{L_1L_2} \times \overrightarrow{L_1L_3}\|} \end{cases} \tag{2}$$

$\mathbf{x}_L$, $\mathbf{y}_L$, and $\mathbf{z}_L$, defined in Eq. (2), can be assembled to form the rotation matrix R_L^D, which represents the rotation from the local coordinate system $\{L\}$ to the ARTS system $\{D\}$:

$$R_L^D = \begin{bmatrix} (\mathbf{x}_L)^T & (\mathbf{y}_L)^T & (\mathbf{z}_L)^T \end{bmatrix} \tag{3}$$

Eq. (3) leads to $\mathbf{T}_L^D$, which is the transformation matrix that includes both rotation and translation components, aligning $\{D\}$ to $\{L\}$:

$$\mathbf{T}_L^D = \begin{bmatrix} R_L^D & (L_1)^T \\ 0 & 1 \end{bmatrix} \tag{4}$$

By inverting Eq. (4), it is possible to calculate $\mathbf{T}_D^L$ to transform the coordinates from the ARTS system $\{D\}$ back to the local system $\{L\}$:

$$\mathbf{T}_D^L = \begin{bmatrix} (R_L^D)^T & -(R_L^D)^T \cdot L_1 \\ 0 & 1 \end{bmatrix} \tag{5}$$

The same reasoning applies to the coordinates of the divot centers obtained from the qualification procedure, referred to measurement reference system $\{Q\}$, introducing the matrix $\mathbf{T}_Q^L$. For the sake of clarity, the letters $\mathbf{p}$ and $\mathbf{q}$ denote the tracked and qualified points, respectively. Using the inverse matrix $\mathbf{T}_D^L$, defined in Eq. (5), and $\mathbf{T}_Q^L$, the equations that transform the point coordinates $\mathbf{p}^D$ and $\mathbf{q}^Q$ from $\{D\}$ and $\{Q\}$, respectively, into their local system $\{L\}$ are given by:

$$\mathbf{p}^L = \mathbf{T}_D^L \cdot \mathbf{p}^D \tag{6}$$

$$\mathbf{q}^L = \mathbf{T}_Q^L \cdot \mathbf{q}^Q \tag{7}$$

Using Eqs. (6) and (7), it is possible to calculate the response variables e_x, e_y, and e_z. These variables represent the differences, i.e. the errors, of the coordinates of points $\mathbf{q}^L$ and $\mathbf{p}^L$ within the local system $\{L\}$:

$$\begin{aligned} e_x &= \mathbf{p}_x^L - \mathbf{q}_x^L \\ e_y &= \mathbf{p}_y^L - \mathbf{q}_y^L \\ e_z &= \mathbf{p}_z^L - \mathbf{q}_z^L \end{aligned} \tag{8}$$

From Eq. (8), the Euclidean distance e_d between $\mathbf{q}^L$ and $\mathbf{p}^L$ is given by:

$$e_d = \sqrt{\left(\mathbf{p}_x^L - \mathbf{q}_x^L\right)^2 + \left(\mathbf{p}_y^L - \mathbf{q}_y^L\right)^2 + \left(\mathbf{p}_z^L - \mathbf{q}_z^L\right)^2} \tag{9}$$

The collected data serve as the basis for the evaluation phase, where statistical analysis and uncertainty estimation are applied under controlled and repeatable conditions. Broader aspects, such as the device-dependent spatial or temporal resolution in highly dynamic scenarios, were not included in the present scope.

3.4. Evaluation

Measured data according to the DOE are analyzed with a Generalized Linear Model (GLM) (Montgomery, 2017; Maculotti et al., 2024) to test for significant factors, while including the 9 measured points per each Geometry level as nested factor (Montgomery, 2017). Stepwise algorithm for term selection is applied, with alpha-to-enter and alpha-to-remove both conventionally set at 15%, to increase degrees of freedom to estimate random errors (Montgomery, 2017; Galetto et al., 2020, 2021).

This works aims at evaluating metrological characteristics of tracking systems, hereby including all relevant influence factors. In particular, accuracy, precision and uncertainty are evaluated. According to the International Vocabulary of Metrology (JCGM, 2008), the accuracy and precision, estimated as reproducibility, can be evaluated as in Eqs. (10) and (11), hereby reported for sake of compactness only for the x-axis, but applicable to quantities of both Eqs. (8) and (9).

$$Acc(\mathbf{p}_x^L) = \overline{e_x} = \frac{\sum_{i=1}^{n} e_{x_i}}{n} \tag{10}$$

$$Repr(\mathbf{p}_x^L) = s(e_x) = \sqrt{\frac{\sum_{i=1}^{n}(e_{x_i} - \overline{e_x})^2}{n-1}} \tag{11}$$

The combined standard uncertainty u, can be evaluated relying on the metrological model (JCGM100, 2008), as per the Procedure for Uncertainty Management (PUMA) method (ISO14253, 2011):

$$\mathbf{p}_x^L = \mathbf{P}_x^L \pm Acc \pm Repr \pm Trac \pm Res \tag{12}$$

where, $\mathbf{P}_x^L$ is the nominal value of the coordinate, and the model includes contributions from accuracy Acc, precision (as reproducibility) $Repr$, traceability $Trac$ and resolution Res. The accuracy can be evaluated as per Eq. (10), and propagating its contribution as a uniform distribution. The reproducibility can be propagated as per its definition in Eq. (11) considering a normal distribution. The traceability term $Trac$ caters for the calibration uncertainty of the qualified reference artifact, i.e., u_{cal}. Finally, the resolution can be propagated as a type-B contribution with a triangular distribution and half range equal to the pixel size pxl of the AR system. Accordingly, it results:

$$\begin{aligned} u^2_{\mathbf{p}_x^L} &= u^2_{Acc} + u^2_{Prec} + u^2_{Trac} + u^2_{Res} \\ u^2_{\mathbf{p}_x^L} &= \frac{\overline{e^2_{x_i}}}{3} + s^2(e_{x_i}) + u^2_{cal_x} + \frac{pxl^2}{6} \end{aligned} \tag{13}$$

The measurement uncertainty U can be evaluated at a conventional 95% confidence interval as:

$$U(\mathbf{p}^L) = t_{1-\frac{1-P}{2},\nu} \cdot u_{\mathbf{p}^L} \tag{14}$$

This formulation allows the quantified uncertainty to be consistently incorporated into the assessment of tooltip positioning accuracy. where $t_{1-\frac{1-P}{2},\nu}$ is the quantile of a t-Student distribution with $\nu = \nu(\mathbf{p}^L)$ degrees of freedom computed according to the Welch-Satterthwaite equation, i.e.:

$$\nu(\mathbf{p}^L) = \frac{u^4_{\mathbf{p}^L}}{\frac{u^4_{Acc}}{n-1} + \frac{u^4_{Prec}}{n-1} + \frac{u^4_{Trac}}{\nu(cal)} + \frac{u^4_{Res}}{100}} \tag{15}$$

where n is the number of data collected and leveraged to evaluate the bias and the precision, the degrees of freedom of the calibration depend on the qualification procedure, in general, and the degrees of freedom for the resolution are set to 100 according to the PUMA method (ISO14253, 2011).

Last, it is worth remarking that the methodology proposed here to estimate measurement uncertainty, caters for systematic measurement errors, by means of the accuracy contribution, which are not corrected but propagated. This is consistent with the great difficulty in correcting, in general, systematic errors that might be sourced by the influence factors described in Fig. 7. Accordingly, the estimation proposed in Eq. (14) when exploiting all the data collected in the experimental plan would estimate the reproduciblity contribution on the total variance of the ANOVA model associated with the GLM, i.e., with $n = 48 \times 9 \times 5 = 2160$ data. This, indeed, will result in a worst case scenario evaluation.

3.5. Case study

Two different ARTSs were tested according to the present protocol.

The first system utilizes the HoloLens 2 as both a tracking and display device. The HoloLens 2 is equipped with several camera sensors: four grayscale visible light cameras, a front-facing RGB camera, and a depth Time of Flight (ToF) camera, which operates in two modes: Hand Tracking and Long Throw Tracking. The tracked tool is equipped with an 80×80 mm ArUco planar marker, as shown in Fig. 9a. The second considered ARTS uses a stereoscopic RGB camera tracking system incorporating two identical Intel RealSense SR305 cameras mounted on a tripod using a custom-made 3D printed frame. The RealSense SR305 cameras were chosen for this setup because they are off-the-shelf devices that can be readily integrated into a custom acquisition system with the use of 3D-printed components. The SR305 has a publicly available datasheet (Intel Corporation, 2019), providing clear documentation for research purposes. Additionally, this camera has been examined in various research studies regarding its applications and metrological performance (Maculotti et al., 2022; Carfagni et al., 2017). These attributes make the SR305 an appropriate option for this system, with well-documented characteristics in the literature. Both the baseline and the camera orientation are adjustable, set according to Fig. 9b. The tracked tool is equipped with four yellow fluorescent spherical markers.

The HMD-based application was developed using the Unity graphics engine, supplemented by the open-source Mixed Reality Toolkit SDK and the Vuforia engine for image target spatial tracking. Due to the different camera sensors, the HoloLens 2 performed Simultaneous Localization and Mapping (SLAM) (Durrant-Whyte and Bailey, 2006; Bailey and Durrant-Whyte, 2006) to establish a fixed and stable reference system $\{D\}$. Consequently, the transformation $\mathbf{T}_H^D$ of the HMD $\{H\}$ was continuously updated, as the HMD was worn by an operator who could move in 6 degrees of freedom (DoF). Simultaneously, the pose $\mathbf{T}_H^T$ of the ArUco marker on the tool, $\{T\}$, was retrieved during the point-picking procedure (Fig. 10a). The tooltip offset was obtained using the scanned mesh. By combining the poses of the HMD and the image target, it was possible to infer the pose of the image target with respect to $\{D\}$, using the following transformation chain:

$$\mathbf{p}^T = \mathbf{T}_H^T \cdot {\mathbf{T}_H^D}^{-1} \cdot \mathbf{p}^D \tag{16}$$

where $\mathbf{p}^T$ is the tooltip offset with respect to $\{T\}$. The point collection is triggered by the air-tap gesture. On the other hand, the spatially fixed position of the RealSense stereoscopic system eliminates the need to track the system itself. The tracking method was implemented in Python version 3.8.10 using the `pyrealsense` and `OpenCV` libraries. The stereo vision algorithm requires a preliminary calibration procedure, performed using the MATLAB Stereo Calibration Toolbox (MathWorks, 2024b), by capturing multiple images of a chessboard from various spatial positions and orientations with both

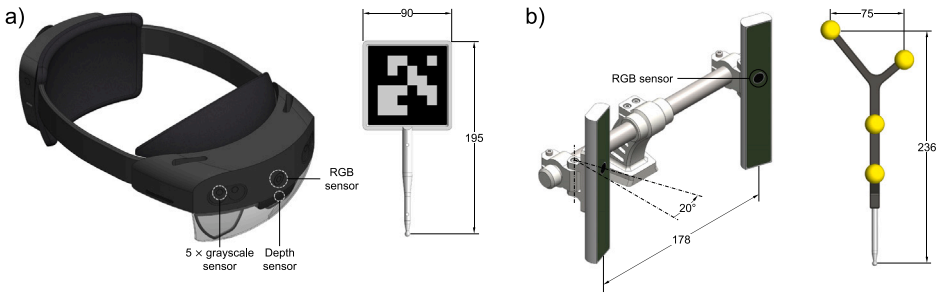

Fig. 9. Tested ARTS: (a) The HMD HoloLens 2, which serves both as a tracking and display system, equipped with an ArUco-based tracked tool; (b) Stereoscopic setup composed of two Intel RealSense SR305 cameras, which track a tool marked with fluorescent spherical markers, paired with a monitor to display augmented information. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cameras. Calibration parameters, such as intrinsic and extrinsic matrices, essential matrix, and distortion coefficients, are calculated and saved into a text file, so that these values are loaded at the start of the algorithm. Before initiating the tracking procedure, a thresholding process is carried out. During this phase, the tool is moved within the field of view of both cameras, and the values for Hue, Saturation, and Value (HSV) are manually adjusted to exclude everything from the image that does not match the color of the markers. This results in a binary mask that isolates only the markers, which appear as near-circular blobs corresponding to the spherical markers on the tool. Subsequently, for each frame, a minimum enclosing circle algorithm is applied to detect blob contours and their centers in u, v coordinates for both the left and right camera frames. Epipolar lines are then calculated to find stereo matching pairs of detected centers between the left and right images. More specifically, epipolar lines are computed using the epipolar constraint (Castillo and Jacobs, 2009), the transformation matrix of the right camera relative to the left one, and the perspective projection equations (Whitehead and Roth, 2004), both obtained from stereo calibration. For a point $\mathbf{x}_l = [u_l, v_l, 1]^T$ in the left image, the corresponding epipolar line in the right image is found by applying the epipolar constraint:

$$\mathbf{x}_r^T \cdot \mathbf{F} \cdot \mathbf{x}_l = 0 \tag{17}$$

Expanding this equation gives:

$$(f_{11}u_l + f_{21}v_l + f_{31})u_r + (f_{12}u_l + f_{22}v_l + f_{32})v_r + (f_{13}u_l + f_{23}v_l + f_{33}) = 0 \tag{18}$$

This defines the epipolar line in the right image as:

$$a_l u_r + b_l v_r + c_l = 0 \tag{19}$$

where a_l, b_l, and c_l are the coefficients of the epipolar line, which depend on the point $\mathbf{x}_l$ in the left image and the fundamental matrix $\mathbf{F}$ which is defined as follows:

$$\mathbf{F} = \mathbf{K}_2^{-T} \cdot \mathbf{E} \cdot \mathbf{K}_1^{-1} \tag{20}$$

$\mathbf{K}_l$ and $\mathbf{K}_r$ are the intrinsic parameter matrices of the left and right cameras, respectively, and $\mathbf{E}$ is the essential matrix (Longuet-Higgins, 1981):

$$\mathbf{E} = [\mathbf{t}]_\times \cdot \mathbf{R}, \quad [\mathbf{t}]_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \tag{21}$$

where $[\mathbf{t}]_\times$ is a skew-symmetric matrix containing the coefficients of the translation vector $\mathbf{t}$, and $\mathbf{R}$ is the rotation matrix between the two cameras. For each marker detected in the left image, the corresponding marker in the right image is found by searching along the epipolar line for the point with the smallest point-to-line distance. The 3D coordinates of each marker's center are calculated using both the imaging equations and the transformation matrix of the right camera to the left one. The projection of a 3D point onto the left and right camera frames is described by the imaging equations:

$$\mathbf{u}_l = \mathbf{P}_l \cdot \mathbf{x}_l, \quad \mathbf{u}_r = \mathbf{P}_r \cdot \mathbf{x}_r \tag{22}$$

where $\mathbf{u}_l$ and $\mathbf{u}_r$ are the 2D points in the left and right images, $\mathbf{P}_l$ and $\mathbf{P}_r$ are the projection matrices for the left and right cameras derived from the intrinsic camera parameters, and $\mathbf{x}_l$ and $\mathbf{x}_r$ are the 3D coordinates of the point in the respective camera frames. The relationship between the 3D coordinates in the two camera frames is given by the transformation matrix $\mathbf{T}$, which includes both rotation $\mathbf{R}$ and translation $\mathbf{t}$:

$$\mathbf{x}_l = \mathbf{T} \cdot \mathbf{x}_r = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \cdot \mathbf{x}_r \tag{23}$$

Eqs. (22) and (23) are combined to form a linear system for solving the unknown 3D coordinates $\mathbf{x}_r = [x_r, y_r, z_r]^T$. This system is written as:

$$\mathbf{A} \cdot \mathbf{x}_r = \mathbf{b} \tag{24}$$

The least squares solution of the system provides the 3D coordinates of the point:

$$\mathbf{x}_r = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \tag{25}$$

This process is repeated for each frame and for all detected blobs, following the increasing order of the v coordinates. Once the 3D coordinates of the detected marker centers are determined relative to the left camera, the ICP algorithm is applied to align the detected 3D marker positions with the known 3D coordinates of the calibrated tool. For each frame, the ICP alignment process starts by generating all possible permutations of the markers to consider any potential unknown order among them. For each permutation, a null initial coarse registration (Genta and Maculotti, 2021) is applied, represented by an identity transformation matrix. The algorithm then iteratively minimizes the alignment error by adjusting the rotation and translation parameters, progressively refining the transformation with each iteration. This refinement continues until the transformation stabilizes, meaning the differences between consecutive iterations fall below a predefined threshold. Once stabilized, the Root Mean Square Error (RMSE) between the transformed model markers and the tracked markers is calculated for each permutation. Finally, the transformation associated with the permutation that yields the lowest RMSE is selected as the optimal alignment for that frame. Algorithm 1 illustrates the stereo-based tracking routine.

Subsequently, the pose estimation of the tool $\mathbf{T}_T^D$ was continuously calculated, with respect to the left camera system $\{D\}$, as shown in Fig. 10b:

$$\mathbf{p}^D = \mathbf{T}_T^D \cdot \mathbf{p}^T \tag{26}$$

Point collection was triggered by pressing a keyboard key.

All custom parts were designed with SolidWorks CAD software and fabricated using a Prusa MK3S machine with white PETG filament from TreeD Filament supplier. For all parts, the same process parameters were employed: a layer thickness of 0.15 mm, 3 surrounding perimeters, 40% fill density, and a gyroid infill pattern. The following section presents the outcome of the experimental procedure outlined above.

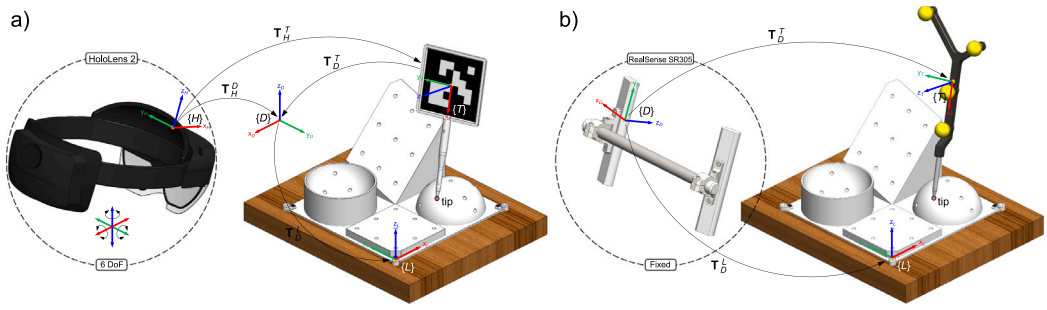

Fig. 10. Transformations involved in the point picking procedure. (a) Hololes 2 $\{H\}$, fixed reference system $\{D\}$, ArUco tool reference system $\{T\}$, and local reference $\{L\}$. (b) RealSense SR305 stereoscopic setup $\{D\}$, spherical marker tool reference system $\{T\}$, and local reference $\{L\}$.

Algorithm 1: Stereo tracking algorithm

Input: Calibration parameters
Output: 3D positions, tool pose
Step 1
HSV thresholding;
Step 2
while $f_l \wedge f_r$ **do**
  $C_l \leftarrow$ `minEnclosingCircle`(f_l);
  $C_r \leftarrow$ `minEnclosingCircle`(f_r);
  for $c_l \in C_l$ **do**
    compute epipolar line e_r;
    $d_{min} \leftarrow \infty$;
    for $c_r \in C_r$ **do**
      $d \leftarrow dist(c_r, e_r)$;
      if $d < d_{min}$ **then**
        $d_{min} \leftarrow d$;
        $best\_match \leftarrow c_r$;
      end
    end
    $M_{tracked} \leftarrow$ `Comp3Dcoord`$(c_l, best\_match)$;
  end
  $P =$ `permute`$(M_{tracked})$;
  for $p \in P$ **do**
    $\mathbf{T}_0 = \mathbf{I}$;
    while $\|\mathbf{T}_{k+1} - \mathbf{T}_k\| \geq \epsilon$ **do**
      $(\mathbf{T}_k) \leftarrow$ `refine`$(\mathbf{T}_k)$;
      err[$\mathbf{T}$] $\leftarrow$ `computeRMSE`$(\mathbf{T}_k, M_{model}, M_{tracked})$;
    end
  end
  $\mathbf{T}^* \leftarrow \arg\min_T$ err[$\mathbf{T}$];
  Estimate tool pose;
  Display frame and tip coordinates
end

4. Results

The calibration artifact, and the tracking tools, were manufactured and qualified by means of the ATOS ScanBox 4105, as shown in Fig. 11.

The DOE described in Section 3.3, was implemented and data analysis carried out by means on MINITAB commercial software, according to the GLM scheme.

Table 2 summarizes the descriptive statistics for the error components e_x, e_y, e_z, and e_d across the factors *Geometry* (G) and *Tool-device distance* (T).

Errors are tendentially larger and more dispersed for the Planar 2, i.e. the tilted plane, than for the Planar 1, i.e. the horizontal plane, but for the x direction. Similar result can be seen comparing the Convex and the Concave shape. Also, the minimum distance seems to worsen errros.

At Min distance for the same geometry, i.e. spherical convex, the HoloLens 2 shows, for response e_z, −0.584 mm with a range of 16.120 mm, compared to the RealSense with 1.247 mm and 11.480 mm respectively.

In the following, results will show details for the e_x, i.e., x-axis, details of other coordinate axes and Cartesian distance d are reported in Appendix A (Fig. A.15 and Fig. A.16) for sake of readability.

Fig. 12 shows the main effect plots, which qualitatively highlights a relevant, i.e., a systematic, contribution from all the considered factors on the measurement error. In addition to the main effects, interactions between the input factors were also analyzed to explore potential combined influences. Fig. 13 shows the interaction plot from which, at a qualitative level, several interactions between considered factors can be considered relevant, e.g., between the *tool-device distance* and the *Operator*.

Fig. 14 shows the histograms of the errors on the Cartesian coordinates for the two considered ARTS. It can be appreciated that errors performed by the HoloLens 2 are less dispersed and more homogeneous in the three Cartesian axes.

Then, formal investigation of influence factors is carried out by means of Analysis of Variance (ANOVA) and GLM. Appendix A reports ANOVA tables for the considered quantities (Tables A.8–A.14). Here, results are summarized, highlighting in Table 3 factors that, with a risk of error of 5%, which systematically affect the responses, i.e. the errors. The results provide a formal support to the hypotheses reported in the literature suggesting the relevant effect of the considered factors. In particular, while all second order interactions are significant, it is worth noticing that very few higher order interaction are; among these a relevant effect is provided by the interaction of Geometry (G), Rotation (R) with other factors, indicating a strong effect (as large as 20% of the total Sum of Squares, see ANOVA table in Appendix A) induced by the spatial orientation with respect to picking direction. An interesting difference can be highlighted between the two considered ARTS devices, i.e. the HoloLense 2 and the RealSense. In particular, the RealSense is characterized by a larger dispersion which is more significantly affected by part positioning; conversely, the HoloLense 2, being less sensitive to the Geometry, the Tool-device distance and the Rotation of the part, enhances the Operator effects.

Once estimated relevant factors, it is possible to address metrological performances evaluation, as per Eq. (13). Table 4 shows mean and standard deviations of errors, i.e., respectively, the accuracy and reproducibility, of the two considered ARTS. As it can be appreciated, consistently with histograms of Fig. 14 RealSense shows worse metrological performances. The accuracy and reproducibility are then combined with the traceability and the resolution to evaluate the measurement uncertainty, reported in Table 5.

5. Discussion

This study proposes applying a procedure based on the traditional metrological approach to investigate the tooltip tracking process for

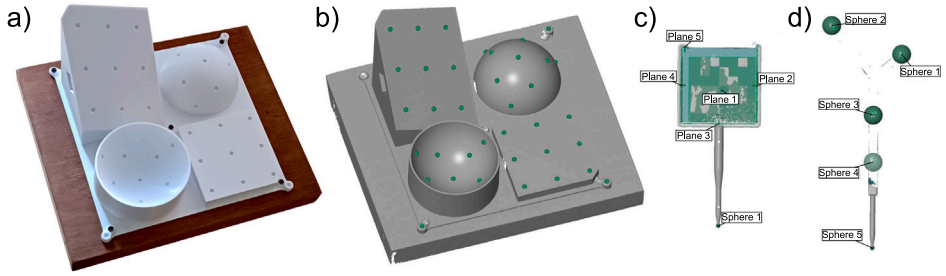

Fig. 11. (a) 3D printed artifact fixed on a wooden base; (b) 3D scanned model of the artifact with green fitting spheres on divots; (c) ArUco 3D scanned model with green fitting planes and tip fitting sphere; (d) Spherical marker tool with green fitting spheres. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Descriptive statistics mean (m), range (R) and standard deviation (s) of errors e for the factors Geometry (G) and Tool-device distance (T), the responses (e) and the ARTSs.

G	T	e	HoloLens 2			RealSense		
			m/mm	R/mm	s/mm	m/mm	R/mm	s/mm
Planar 1	Max	e_x	−0.139	15.510	1.362	−0.596	10.940	1.505
	Min		−0.461	8.190	1.537	−0.526	7.720	1.511
	Max	e_y	−0.253	15.230	1.506	−0.001	14.150	2.575
	Min		0.003	21.110	2.510	−0.647	11.300	2.000
	Max	e_z	−0.242	8.380	0.910	0.671	9.990	1.522
	Min		−0.106	11.890	1.323	0.543	8.140	1.484
	Max	e_d	1.794	11.030	1.366	2.909	9.540	1.879
	Min		2.600	11.440	1.961	2.772	7.800	1.330
Planar 2	Max	e_x	−0.082	11.420	1.873	−1.574	18.220	3.047
	Min		−0.490	15.050	2.529	−2.452	10.280	2.332
	Max	e_y	−0.071	25.620	2.221	0.536	23.920	3.640
	Min		0.155	15.640	2.064	−3.492	13.550	2.972
	Max	e_z	0.427	9.750	1.470	1.794	13.690	2.170
	Min		−0.128	9.560	1.393	2.604	8.110	1.613
	Max	e_d	2.651	14.100	1.935	5.007	14.070	2.844
	Min		3.033	9.430	1.909	6.200	9.270	1.836
Spherical concave	Max	e_x	−0.237	11.480	1.903	−1.788	15.370	2.808
	Min		−0.839	16.110	2.556	−2.046	10.660	2.632
	Max	e_y	−0.103	12.930	1.963	0.065	20.430	3.111
	Min		−0.104	17.440	2.669	−3.562	15.380	2.215
	Max	e_z	−0.379	8.830	1.307	1.396	10.950	1.988
	Min		−0.687	8.600	1.408	1.289	10.790	1.782
	Max	e_d	2.622	8.460	1.581	4.494	12.530	2.531
	Min		3.514	10.470	2.107	5.30	12.880	2.327
Spherical convex	Max	e_x	−0.013	14.820	1.528	−0.365	10.150	1.530
	Min		−0.207	14.260	1.913	−1.168	8.850	1.610
	Max	e_y	−0.323	20.980	2.245	−0.236	20.740	2.916
	Min		0.047	22.490	2.639	−0.786	11.120	2.145
	Max	e_z	−0.073	16.830	1.649	0.594	13.590	1.828
	Min		−0.584	16.120	1.804	1.247	11.480	1.732
	Max	e_d	2.463	14.860	2.029	3.227	14.440	2.069
	Min		3.012	15.370	2.270	3.406	7.860	1.448

ARTS. Additionally, a specific measuring artifact was proposed to collect data points. The procedure involves qualifying both the artifact and the tracked tool, the Design of Experiments (DOE), and assessing metrological performance. Moreover it involves both accuracy and reproducibility calculation and the evaluation of the measurements uncertainty, at a conventional 95% confidence level. The entire procedure is applied to two different ARTS: the HoloLens 2 HMD combined with a planar ArUco-equipped tracked tool, and a monitor-based application that includes a stereoscopic RealSense SR305 camera setup together with a tool featuring yellow fluorescent spherical markers.

The results shown in Tables 4 and 5 indicated that the HoloLens 2 generally outperformed the Intel RealSense stereo system in terms of metrological performance. The accuracy values achieved using the HoloLens 2 are −0.309 mm, −0.081 mm, and −0.222 mm along the x, y, and z axes, respectively. In terms of uncertainty, the results showed figures of 3.988 mm, 4.563 mm and 3.009 mm for the three axes.

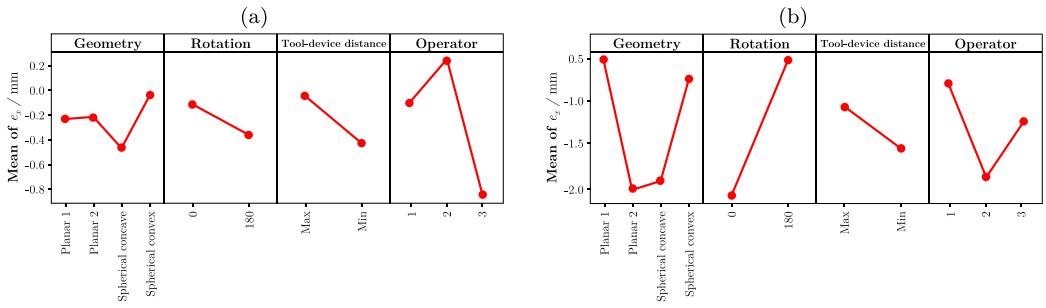

Fig. 12. Main effect plots for e_x due to (a) HoloLens 2 and (b) Intel RealSense.

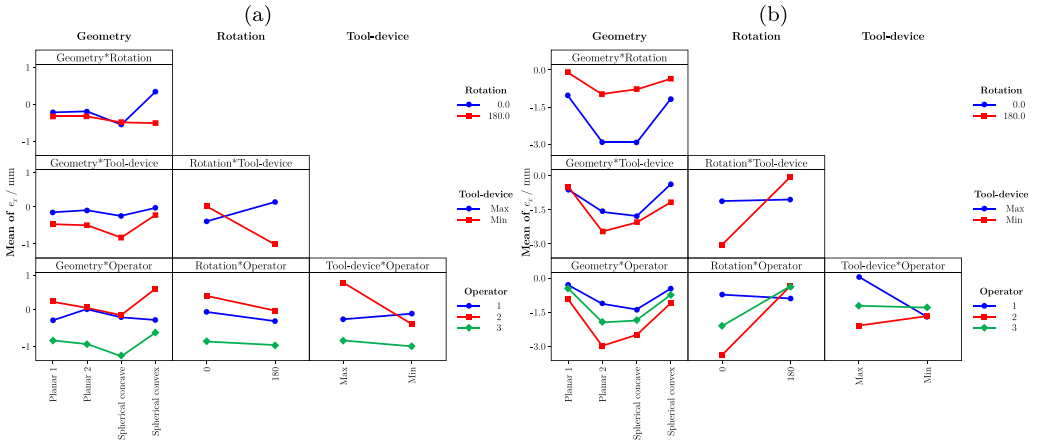

Fig. 13. Interaction plots for e_x due to (a) HoloLens 2 and (b) Intel RealSense.

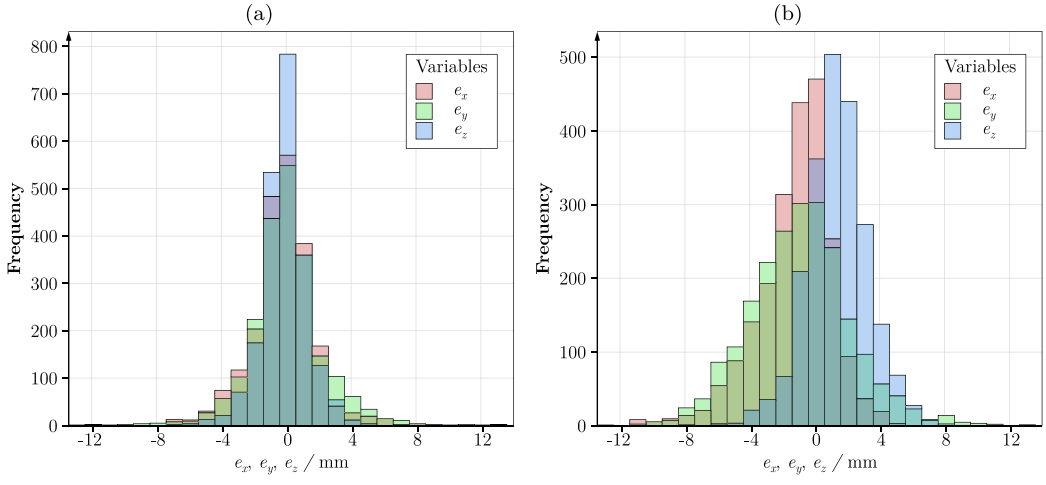

Fig. 14. Grouped histograms of response variables e_x, e_y and e_z for (a) HoloLens 2 and (b) Intel RealSense.

In contrast, the accuracy values obtained with the RealSense stereo setup are −1.314 mm, −1.015 mm, and 1.267 mm with measurement uncertainties of 4.981 mm, 6.370 mm and 4.067 mm. These result in a Cartesian distance accuracy of 2.711 mm for the HoloLens 2 compared to 4.164 mm for the RealSense, with a measurement uncertainty of 5.071 mm and 6.800 mm, respectively. Overall, the HoloLens

Table 3
p-values of the considered factors on the error responses. Factors: Geometry (G), Rotation (R), Tool-device distance (D), Operator (O), Measurement point (P) nested in Geometry. Only significant factor (with a risk of error of 0.05) are reported.

Effects	HoloLens 2			RealSense		
	e_x	e_y	e_z	e_x	e_y	e_z
G	0.003	***	<0.001	<0.001	<0.001	<0.001
R	0.002	<0.001	0.011	<0.001	<0.001	<0.001
D	<0.001	0.024	<0.001	<0.001	<0.001	<0.001
O	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
G*R	<0.001	<0.001	***	<0.001	<0.001	0.009
G*D	**	***	<0.001	<0.001	<0.001	<0.001
G*O	0.024	0.002	<0.001	<0.001	**	<0.001
P(G)	***	***	<0.001	<0.001	0.003	<0.001
R*D	<0.001	***	*	<0.001	0.004	<0.001
R*O	**	<0.001	***	<0.001	**	<0.001
D*O	<0.001	***	<0.001	<0.001	<0.001	<0.001
G*R*D	0.015	<0.001	**	<0.001	<0.001	<0.001
G*R*O	0.017	0.004	0.001	<0.001	<0.001	<0.001
R*P(G)	***	**	***	***	<0.001	**
G*D*O	***	***	0.001	0.032	0.024	0.014
D*P(G)	***	***	***	***	<0.001	***
O*P(G)	***	***	0.023	***	***	***
R*D*O	<0.001	***	<0.001	<0.001	0.009	0.002
R*D*P(G)	***	***	***	***	***	**
R*O*P(G)	***	***	***	***	***	***
D*O*P(G)	***	***	***	***	***	***

Non-significant factors are synthesized:
* (0.05 < p-value < 0.1).
** (0.1 < p-value < 0.35).
*** (p-value > 0.35).

Table 4
Accuracy and reproducibility for HoloLense 2 and RealSense.

	HoloLens 2		RealSense	
	Acc/mm	*Repr*/mm	*Acc*/mm	*Repr*/mm
e_x	−0.3086	1.9589	−1.3143	2.3204
e_y	−0.0811	2.2577	−1.0154	3.1273
e_z	−0.2215	1.4633	1.2672	1.8917
e_d	2.7111	1.9680	4.1643	2.3998

2 demonstrated better accuracy in tooltip tracking compared to the Intel RealSense. The comparison of the obtained accuracy with the measurement uncertainty, and also with the reproducibility, shows that systematic errors cannot be considered statistically significant due to the relatively large dispersion. The obtained results also highlighted the influence of various factors on the measurement errors. For instance, the interaction between tool-device distance and operator significantly impacted the measurement accuracy (see e.g. Fig. 13 and Table A.7). These findings underscore the importance of optimizing the tool-device distance to reduce errors, which can be achieved by maintaining an optimal distance that avoids distortions caused by lens aberrations, for example.

The obtained results allow for estimating the accuracy and reproducibility of an ARTS and investigating its suitability for specific applications by identifying relevant influencing factors and calculating the measurement uncertainty. Moreover, the DOE approach helps optimize measurement conditions to reduce uncertainty. As mentioned in Section 3.4, such uncertainty propagation, by leveraging the total variance of collected data, which thus includes contributions due to systematic factors, is a worst case scenario.

Accordingly, it can be of interest attempting a management of the systematic errors to reduce errors and the resulting uncertainty. In particular, the results of the GLM shows, see Fig. 12, that errors can be reduced by increasing the distance between the tool and the device. This is consistent with a better imaging of the measurand at the center of the field of view, which allows avoiding distortions introduced by lens aberrations at the edge (that are relevant at a smaller distance which images the object in the full field of view). Similarly, given the substantially different measurement tasks associated with diverse geometrical features, a task-specific measurement uncertainty can be reported per each geometry, i.e., horizontal plane, tilted plane, concave and convex sphere. Conversely, operators and relative rotations cannot in general be controlled, and thus the related effects reduced. Table 6 shows the results of such procedure to reduce the uncertainty while specifying it for feature-specific tracking, which shows a definitive improvement of performances.

In particular, the accuracy contribution is generally highly reduced. Furthermore, it can be appreciated that most critical features are tilted plane and convex spheres.

Last, it is interesting to benchmark the metrological performances of the two considered ARTS.

First, accuracy can be compared. As shown in both Tables 5 and 6, the HoloLense 2 has systematic errors smaller of up to one order of magnitude than the RealSense.

Then, a more aggregated analysis can be addressed by comparing standard combined variances u^2 by means of a hypothesis test based on Fisher distribution. In particular, a monolateral test, at a confidence level of 95% was carried out with null hypothesis $H_0 : u^2_{e,H}/u^2_{e,R} = 1$ tested against an alternative hypothesis $H_1 : u^2_{e,H}/u^2_{e,R} < 1$, where the subscript $_H$ and $_R$ indicates the HoloLense 2 and the Realsens, respectively, and the test is applied to all evaluated errors e, i.e., e_x, e_y, e_z and e_d both in case of worst case scenario, as reported in Table 5, and considering errors reduction, as in Table 6. In all cases the null hypothesis was rejected with a risk of error of 5% to favor the alternative hypothesis. Accordingly, the HoloLense 2 can be considered having better metrological performances.

Expanded uncertainty estimation allows for assessing whether the measurements performed by the tested ARTS meet the tolerances required for specific applications. In the following, the expanded uncertainty and tolerance specification limits are compared based on ISO 14253-1:2017 (ISO14253, 2017): a positive acceptance zone indicates that the use of the specific ARTS conforms to the application's requirements. In surgical domain, Cartiaux et al. (2008) defined a safe specification zone width (Kholkhujaev et al., 2022) of 10 mm in

Table 5
Uncertainty (with $k = 2$) and uncertainty contributors for HoloLens 2 and RealSense (worst case scenario).

	HoloLens 2						RealSense					
	u_{Acc}/mm	u_{Prec}/mm	u_{Trac}/mm	u_{Res}/mm	u/mm	U/mm	u_{Acc}/mm	u_{Prec}/mm	u_{Trac}/mm	u_{Res}/mm	u/mm	U/mm
e_x	0.178	1.959	0.002	0.327	1.994	3.988	0.759	2.320	0.002	0.143	2.446	4.891
e_y	0.047	2.258			2.282	4.563	0.586	3.127			3.185	6.370
e_z	0.128	1.463			1.505	3.009	0.732	1.892			2.033	4.067
e_d	1.565	1.968			2.536	5.071	2.404	2.400			3.400	6.800

Table 6
Task-based uncertainty contribution and expanded uncertainty (with $k = 2$) for HoloLens 2 and RealSense at a maximum tool-device distance for different Geometeries (G).

G	V	HoloLens 2				RealSense			
		u_{Acc}/mm	u_{Prec}/mm	u/mm	U/mm	u_{Acc}/mm	u_{Prec}/mm	u/mm	U/mm
Planar 1	e_x	0.080	1.362	1.403	2.805	0.344	1.505	1.578	3.155
	e_y	0.146	1.506	1.548	3.096	0.001	2.575	2.596	5.191
	e_z	0.140	0.910	0.977	1.954	0.387	1.522	1.604	3.208
	e_d	1.035	1.366	1.745	3.489	1.680	1.879	2.541	5.083
Planar 2	e_x	0.047	1.873	1.902	3.804	0.909	3.047	3.196	6.393
	e_y	0.041	2.221	2.245	4.491	0.309	3.640	3.668	7.335
	e_z	0.246	1.470	1.525	3.051	1.036	2.170	2.427	4.853
	e_d	1.531	1.935	2.489	4.977	2.891	2.844	4.068	8.137
Spherical concave	e_x	0.137	1.903	1.936	3.871	1.032	2.808	3.010	6.019
	e_y	0.059	1.963	1.991	3.982	0.038	3.111	3.128	6.257
	e_z	0.219	1.307	1.365	2.729	0.806	1.988	2.170	4.340
	e_d	1.514	1.581	2.213	4.425	2.595	2.531	3.639	7.279
Spherical convex	e_x	0.008	1.528	1.563	3.126	0.211	1.530	1.578	3.156
	e_y	0.186	2.245	2.276	4.553	0.136	2.916	2.937	5.875
	e_z	0.042	1.649	1.682	3.363	0.343	1.828	1.888	3.777
	e_d	1.422	2.029	2.499	4.998	1.863	2.069	2.803	5.607

osseous pelvic tumors resection. It could be reasonably refer red to the tilted planar geometry, i.e., the Planar 2 geometry, as per the planar geometry osteotomies traditional planar shape (Pietruski et al., 2019; Zhou et al., 2017; Gao et al., 2019). Accordingly, a standard uncertainty u on the cartesian distance, according to Table 6, 2.489 mm and 4.068 mm for HoloLens 2 and RealSense stereo respectively can be considered. This results in the HoloLens meeting the tolerance requirements, whilst the use of the RealSense stereo system is not recommended. For another biomedical application, Rampersaud et al. (2001) defined a tolerance value for spinal pedicle screw placement using image-guided techniques by developing a geometric model where both the pedicle and the screw are modeled as cylinders of larger and smaller diameters, respectively. For the lumbar vertebra L5, the maximum permissible error tolerances are 3.8 mm and 12.7°, resulting in a negative acceptance zone for both ARTS, making their use in this specific application not advisable.

Bräuer et al. (2003) assessed the precision requirements for the fabrication and assembly of the Wendelstein 7-X (W7-X) stellarator, a nuclear fusion reactor, to ensure the high geometrical accuracy of the magnet system assembly. They determined that a maximum positional error of 10 mm is permissible for one of the 50 non-planar coils. A standard uncertainty u in the Cartesian distance of 2.489 mm for the HoloLens 2 and 4.068 mm for the RealSense stereo system, as shown in Table 6, can be considered. Consequently, the HoloLens 2 meets the tolerance requirements, while the RealSense stereo system does not meet these specifications.

Overall, this study applies a metrological approach to evaluate the tracking performance of AR systems, which can also be extended to any optical tracking system. Assessing optical tracking systems presents a complex set of challenges due to variations in hardware, software, and environmental dependencies. These inconsistencies make it difficult to perform direct comparisons or establish standardized benchmarks. Consequently, there is a critical need for a methodology that can objectively assess and compare the performance of heterogeneous optical tracking systems across various operational scenarios. For this reason, the key contribution of the current work is enabling a systematic analysis of key performance metrics such as accuracy, precision, and uncertainty. Such a methodology facilitates more transparent evaluations and enables researchers and practitioners to identify the most suitable systems for their applications, while also driving improvements in tracking technology through more robust and consistent performance metrics. Moreover, it is adaptable for optimizing system parameters, testing new AR tracking technologies, or benchmarking different tracking systems, providing a practical tool for assessing performance across a variety of applications. To validate the approach, two tracking systems of different nature were tested: the HoloLens 2, which uses a mobile headset with marker-based tracking, and the RealSense system, which relies on fixed cameras and a different marker detection method. Despite these differences in hardware setup, marker types, and tracking algorithms, the methodology enabled a meaningful comparison through quantitative analysis. This allowed for a reliable performance assessment, demonstrating the robustness of the approach in evaluating AR tracking systems with different configurations.

6. Conclusions

This work presents a standardized metrological methodology for evaluating the tracking performance of ARTS, which is also applicable to other optical tracking systems requiring tool tracking. The methodology systematically assesses key metrics such as accuracy, precision, and expanded uncertainty, while identifying significant factors that influence tracking error. A framework for designing a tailored DOE was developed, allowing users to customize the evaluation process for specific applications. Additionally, an ad hoc artifact was introduced to measure points on various surface geometries, enabling a comprehensive analysis of digitization performance. The proposed method supports both performance evaluation and parameter optimization, as well as benchmarking different systems, providing a valuable tool

for system assessment and optimization across various applications. The effectiveness of the proposed approach was demonstrated by testing two low-cost, state-of-the-art tracking systems: HoloLens 2 and a stereoscopic system with dual RealSense cameras. The HoloLens 2 exhibited superior accuracy and precision under the tested conditions. Although there were notable differences in hardware, such as the use of mobile versus fixed cameras, as well as variations in markers and tracking algorithms, the methodology still enabled an effective comparison between the systems. The quantitative analysis offered a reliable evaluation, emphasizing the adaptability and strength of the proposed approach in handling diverse tracking configurations. In future work, the proposed methodology will be used to optimize tracking system design, focusing on software and marker alternatives with consistent hardware configurations. At the same time, the framework could be extended to include benchmarks on spatial and temporal resolution in dynamic scenarios, further complementing the accuracy-oriented evaluation presented here.

CRediT authorship contribution statement

Federico Salerno: Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Luca Ulrich:** Writing – original draft, Visualization, Validation, Software, Methodology, Data curation, Conceptualization. **Giacomo Maculotti:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Sandro Moos:** Writing – review & editing, Supervision, Resources, Conceptualization. **Gianfranco Genta:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis. **Enrico Vezzetti:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Maurizio Galetto:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

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.

Acknowledgments

F. Salerno is attending the PhD programme in Gestione, Produzione e Design at Politecnico di Torino, Cycle XXXVIII, with the support of a scholarship co-financed by the Ministerial Decree no. 352 of 9th April 2022, based on the NRRP - funded by the European Union - NextGenerationEU - Mission 4C2, Investment 3.3, and by the Foundation ITS Biotecnologie e Nuove Scienze della Vita - Piemonte.

G. Maculotti, G. Genta and M. Galetto would like to acknowledge funding from the 23IND08 DI-Vision project, which received funding from the European Partnership on Metrology, co-financed by the European Union's Horizon Europe Research and Innovation Programme and by the Participating States.

Appendix A

A.1. Post-processing routine

This work incorporates a post-processing routine, logically represented by Algorithm 2, written in Python to manage all the points $\mathbf{p}^D$ collected in the experiment and provide response variables e_x, e_y, and e_z. The 20 runs of each operator produce 60 runs for each ARTS. The tooltip coordinates of each run are collected in separate .txt files in a specific repository, along with the unique file containing the coordinates of qualified points $\mathbf{q}_Q$. The code takes as input the repository containing the .txt files and the $\mathbf{q}^Q$ file, while the output data includes $\mathbf{T}_D^L$, $\mathbf{T}_Q^L$, $\mathbf{p}L, Q$, $\mathbf{p}^L$, e_x, e_y, and e_z. Three functions were defined to compute $\mathbf{T}_D^L$ (and $\mathbf{T}_Q^L$), the coordinate transformation with respect to L, and the response variables, respectively. A main loop is used to parse the .txt files in the path, compute the response variables for each file, and output them.

Algorithm 2: Response variable calculation e_x, e_y and e_z

```
Data: p^D_file_paths and q^Q_file
Result: T^L_D, T^L_Q, q^L, p^L, e_x, e_y, e_z
function T_calculation(p):
    x_L ← (p[:][1] − p[:][0]) / ‖p[:][1] − p[:][0]‖;
    v ← (p[:][2] − p[:][0]);
    z_L ← (x_L × V) / ‖Z_L‖;
    y_L ← z_L × x_L;
    t ← −p[:][0];
    R ← [x_L  y_L  z_L];
    T^{-1} ← [R  t; 0  1];
    T ← Invert(T^{-1});
    return T;
function coord_transformation(p, T):
    p_T ← [];
    for point ∈ p do
        point_h ← [point; 1];
        point_T ← T · point_h;
        Append to p_T the first three components of point_T;
    end
    return p_T;
function e_calculation(p^L, q^L):
    e ← [];
    for i ← 3 to length(p^L) do
        e_x ← p^L[i][0] − q^L[i][0];
        e_y ← p^L[i][1] − q^L[i][1];
        e_z ← p^L[i][2] − q^L[i][2];
        Append to e the list [e_x, e_y, e_z];
    end
    return e;
q^Q ← read_points(q^Q_file);
T^L_Q ← T_calculation(q^Q);
q^L ← coord_transformation(q^Q, T^L_Q);
file_paths ← search(p^D_file_paths/*.txt);
/* Main processing loop                                   */
for i ∈ file_paths do
    p^D ← read_points(i);
    T^L_P ← T_calculation(p^D);
    p^L ← coord_transformation(p^D, T^L_D);
    e ← e_calculation(p^L, q^L);
    print(e)
end
```

A.2. Graphs

See Figs. A.15 and A.16.

A.3. ANOVA tables

See Tables A.7–A.14.

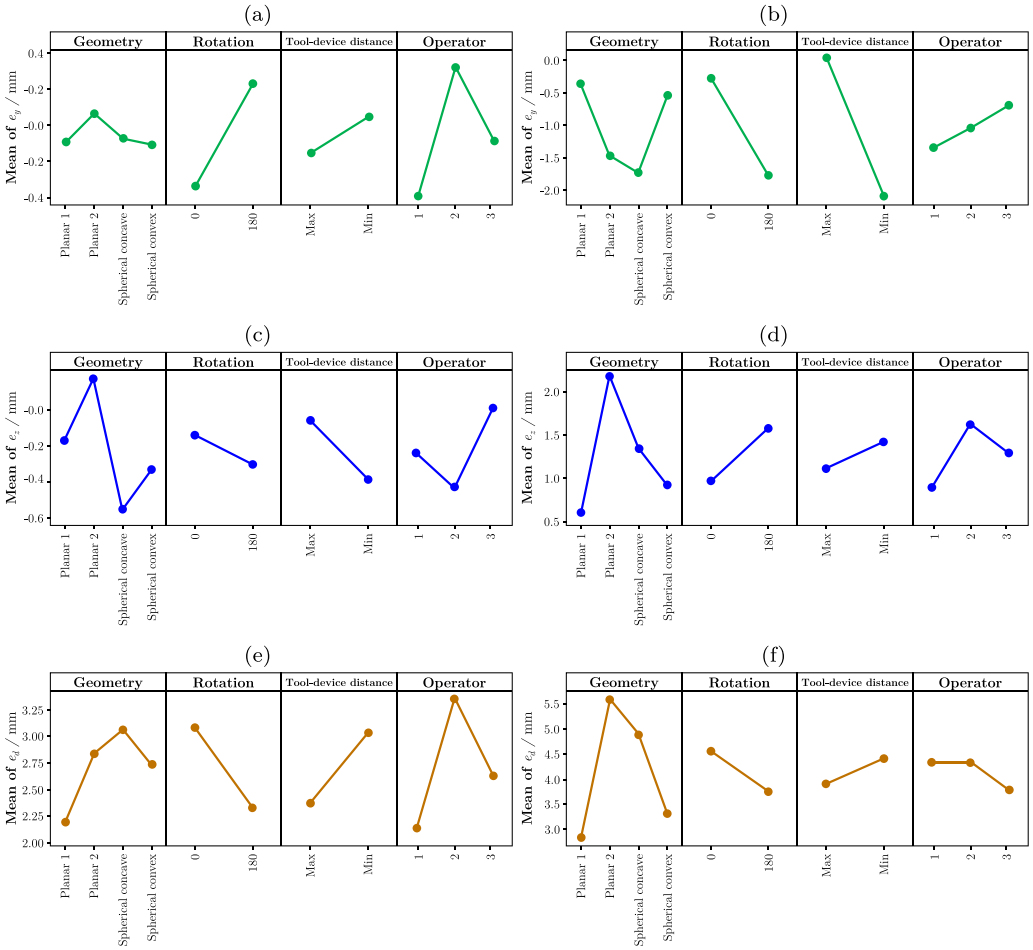

Fig. A.15. Main effect plots for: (a) e_y due to HoloLens 2, (b) e_y due to RealSense, (c) e_z due to HoloLens 2, (d) e_z due to RealSense, (e) e_d due to HoloLens 2, and (f) e_d due to RealSense.

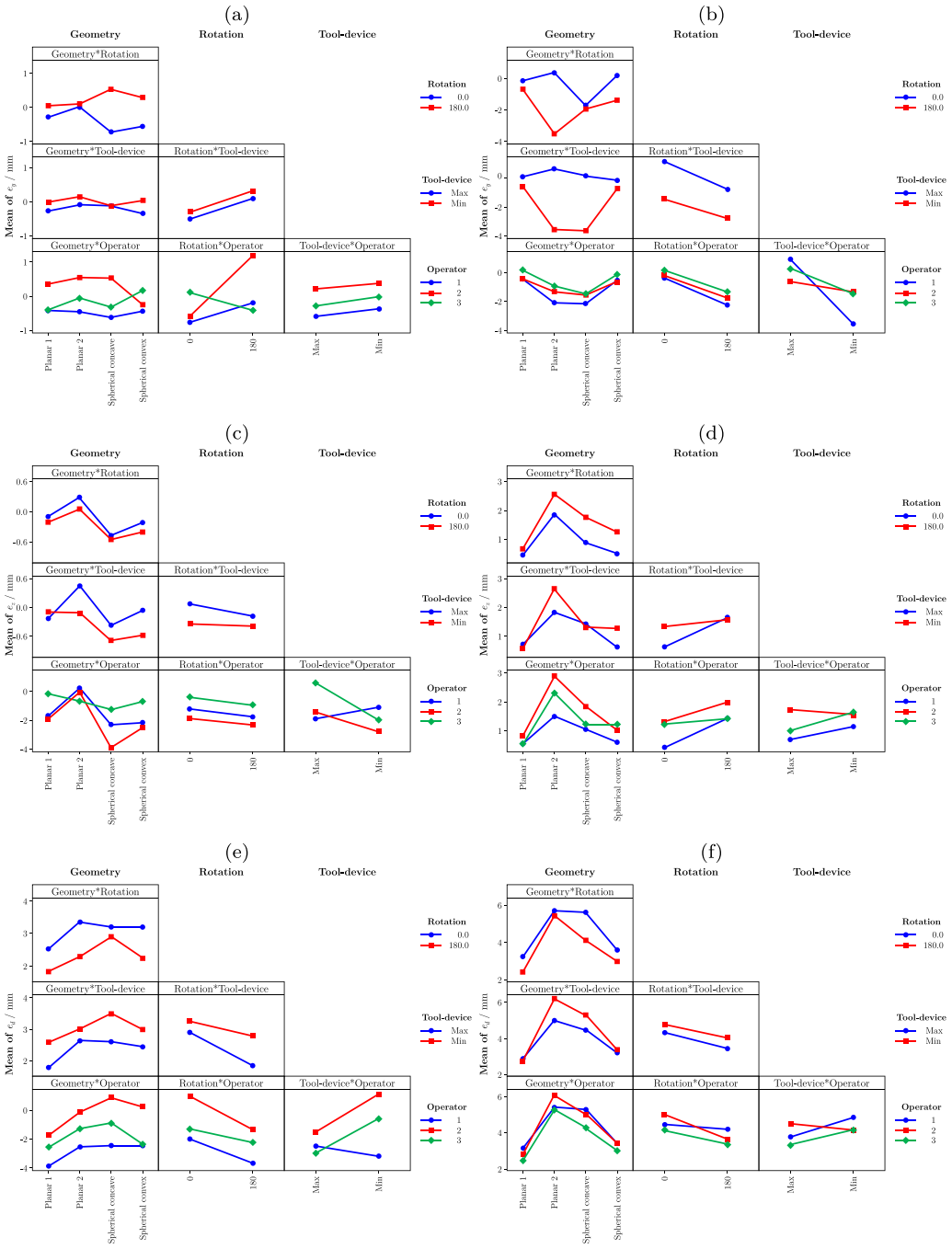

Fig. A.16. Interaction plots for: (a) e_y due to HoloLens 2, (b) e_y due to RealSense, (c) e_z due to HoloLens 2, (d) e_z due to RealSense, (e) e_d due to HoloLens 2, and (f) e_d due to RealSense.

Table A.7
HoloLens 2 e_x, ANOVA.

Source	DF	Adj SS	Adj MS	F-value	P-value
Geometry	3	50.09	16.696	5.30	0.001
Rotation	1	34.45	34.454	10.94	0.001
Tool-device distance	1	78.49	78.494	24.93	0.000
Operator	2	455.29	227.646	72.30	0.000
Geometry*Rotation	3	65.20	21.735	6.90	0.000
Geometry*Operator	6	51.45	8.575	2.72	0.012
Rotation*Tool-device distance	1	344.85	344.848	109.52	0.000
Tool-device distance*Operator	2	164.76	82.380	26.16	0.000
Geometry*Rotation*Tool-device distance	3	37.01	12.336	3.92	0.008
Geometry*Rotation*Operator	6	54.34	9.057	2.88	0.009
Rotation*Tool-device distance*Operator	2	244.69	122.345	38.85	0.000
Error	2129	6703.76	3.149		
Lack-of-Fit	401	491.45	1.226	0.34	1.000
Pure Error	1728	6212.31	3.595		
Total	2159	8284.39			

Table A.8
HoloLens 2 e_y, ANOVA.

Source	DF	Adj SS	Adj MS	F-value	P-value
Rotation	1	193.9	193.872	42.72	0.000
Tool-device distance	1	24.4	24.431	5.38	0.020
Operator	2	210.8	105.408	23.23	0.000
Geometry*Rotation	3	100.3	33.439	7.37	0.000
Geometry*Operator	6	102.0	17.005	3.75	0.001
Rotation*Operator	2	462.4	231.195	50.94	0.000
Geometry*Rotation*Tool-device distance	3	128.1	42.692	9.41	0.000
Geometry*Rotation*Operator	6	93.4	15.559	3.43	0.002
Error	2135	9 689.6	4.538		
Lack-of-Fit	407	1 236.1	3.037	0.62	1.000
Pure Error	1728	8 453.4	4.892		
Total	2159	11 004.8			

Table A.9
HoloLens 2 e_z, ANOVA.

Source	DF	Adj SS	Adj MS	F-value	P-value
Geometry	3	134.28	44.759	24.37	0.000
Rotation	1	12.38	12.382	6.74	0.009
Tool-device distance	1	51.75	51.751	28.17	0.000
Operator	2	62.52	31.259	17.02	0.000
Geometry*Tool-device distance	3	40.30	13.433	7.31	0.000
Geometry*Operator	6	57.92	9.653	5.26	0.000
Point(Geometry)	32	142.67	4.458	2.43	0.000
Rotation*Tool-device distance	1	5.92	5.918	3.22	0.073
Tool-device distance*Operator	2	91.56	45.781	24.92	0.000
Geometry*Rotation*Operator	6	42.92	7.153	3.89	0.001
Geometry*Tool-device distance*Operator	6	42.83	7.139	3.89	0.001
Operator*Point(Geometry)	64	169.78	2.653	1.44	0.013
Rotation*Tool-device distance*Operator	2	39.01	19.507	10.62	0.000
Error	2030	3729.00	1.837		
Lack-of-Fit	302	385.36	1.276	0.66	1.000
Pure Error	1728	3343.65	1.935		
Total	2159	4622.84			

Table A.10
RealSense SR305 e_x, ANOVA.

Source	DF	Adj SS	Adj MS	F-value	P-value
Geometry	3	928.2	309.42	118.96	0.000
Rotation	1	1 261.3	1261.32	484.92	0.000
Tool-device distance	1	117.8	117.84	45.30	0.000
Operator	2	398.8	199.40	76.66	0.000
Geometry*Rotation	3	203.2	67.74	26.04	0.000
Geometry*Tool-device distance	3	82.7	27.57	10.60	0.000
Geometry*Operator	6	95.8	15.97	6.14	0.000
Point(Geometry)	32	227.6	7.11	2.73	0.000
Rotation*Tool-device distance	1	1 145.6	1145.57	440.42	0.000
Rotation*Operator	2	940.5	470.26	180.80	0.000
Tool-device distance*Operator	2	468.7	234.34	90.09	0.000
Geometry*Rotation*Tool-device distance	3	149.4	49.81	19.15	0.000
Geometry*Rotation*Operator	6	86.5	14.42	5.54	0.000
Geometry*Tool-device distance*Operator	6	39.0	6.50	2.50	0.021
Rotation*Tool-device distance*Operator	2	53.5	26.73	10.27	0.000
Error	2086	5 425.8	2.60		
Lack-of-Fit	358	440.2	1.23	0.43	1.000
Pure Error	1728	4 985.6	2.89		
Total	2159	11 624.6			

Table A.11
RealSense SR305 e_y, ANOVA.

Source	DF	Adj SS	Adj MS	F-value	P-value
Geometry	3	801.7	267.23	51.77	0.000
Rotation	1	1 306.6	1306.60	253.11	0.000
Tool-device distance	1	2 644.4	2644.45	512.28	0.000
Operator	2	172.4	86.19	16.70	0.000
Geometry*Rotation	3	1 096.1	365.38	70.78	0.000
Geometry*Tool-device distance	3	1 420.2	473.40	91.71	0.000
Geometry*Operator	6	53.9	8.99	1.74	0.108
Point(Geometry)	32	306.6	9.58	1.86	0.003
Rotation*Tool-device distance	1	43.9	43.91	8.51	0.004
Tool-device distance*Operator	2	1 226.6	613.28	118.80	0.000
Geometry*Rotation*Tool-device distance	3	331.0	110.32	21.37	0.000
Geometry*Rotation*Operator	6	140.1	23.35	4.52	0.000
Rotation*Point(Geometry)	32	583.0	18.22	3.53	0.000
Geometry*Tool-device distance*Operator	6	76.7	12.79	2.48	0.022
Tool-device distance*Point(Geometry)	32	413.2	12.91	2.50	0.000
Rotation*Tool-device distance*Operator	2	50.0	24.98	4.84	0.008
Error	2024	10 448.1	5.16		
Lack-of-Fit	296	1 378.8	4.66	0.89	0.903
Pure Error	1728	9 069.3	5.25		
Total	2159	21 114.5			

Table A.12
RealSense SR305 e_z, ANOVA.

Source	DF	Adj SS	Adj MS	F-value	P-value
Geometry	3	772.70	257.566	96.62	0.000
Rotation	1	207.29	207.291	77.76	0.000
Tool-device distance	1	50.86	50.858	19.08	0.000
Operator	2	189.28	94.640	35.50	0.000
Geometry*Rotation	3	31.44	10.479	3.93	0.008
Geometry*Tool-device distance	3	98.86	32.952	12.36	0.000
Geometry*Operator	6	97.16	16.193	6.07	0.000
Point(Geometry)	32	233.13	7.285	2.73	0.000
Rotation*Tool-device distance	1	80.36	80.358	30.14	0.000
Rotation*Operator	2	59.82	29.912	11.22	0.000
Tool-device distance*Operator	2	64.69	32.346	12.13	0.000
Geometry*Rotation*Tool-device distance	3	74.76	24.919	9.35	0.000
Geometry*Rotation*Operator	6	71.14	11.857	4.45	0.000
Rotation*Point(Geometry)	32	112.30	3.509	1.32	0.111
Geometry*Tool-device distance*Operator	6	43.31	7.219	2.71	0.013
Rotation*Tool-device distance*Operator	2	34.34	17.168	6.44	0.002
Rotation*Tool-device distance*Point(Geometry)	32	114.36	3.574	1.34	0.097
Error	2022	5390.19	2.666		
Lack-of-Fit	294	758.44	2.580	0.96	0.658
Pure Error	1728	4631.75	2.680		
Total	2159	7725.97			

Table A.13
HoloLens 2 e_d, ANOVA.

Source	DF	Adj SS	Adj MS	F-value	P-value
Geometry	3	221.18	73.726	25.47	0.000
Rotation	1	309.64	309.643	106.96	0.000
Tool-device distance	1	233.70	233.695	80.72	0.000
Operator	2	543.94	271.969	93.94	0.000
Geometry*Rotation	3	46.00	15.333	5.30	0.001
Geometry*Tool-device distance	3	22.25	7.418	2.56	0.053
Geometry*Operator	6	48.78	8.131	2.81	0.010
Point(Geometry)	32	202.47	6.327	2.19	0.000
Rotation*Tool-device distance	1	48.62	48.618	16.79	0.000
Rotation*Operator	2	38.17	19.087	6.59	0.001
Tool-device distance*Operator	2	264.78	132.389	45.73	0.000
Geometry*Rotation*Operator	6	162.76	27.127	9.37	0.000
Geometry*Tool-device distance*Operator	6	40.52	6.753	2.33	0.030
Rotation*Tool-device distance*Operator	2	131.08	65.530	22.64	0.000
Error	2089	6047.81	2.895		
Lack-of-Fit	361	715.04	1.981	0.64	1.000
Pure Error	1728	5332.78	3.086		
Total	2159	8361.68			

Table A.14
RealSense e_d, ANOVA.

Source	DF	Adj SS	Adj MS	F-value	P-value
Geometry	3	2743.2	914.415	261.59	0.000
Rotation	1	353.35	353.347	101.78	0.000
Tool-device distance	1	140.64	140.633	40.23	0.000
Operator	2	113.8	56.899	16.30	0.000
Geometry*Rotation	3	112.7	37.579	10.75	0.000
Geometry*Tool-device distance	3	49.87	16.622	4.75	0.003
Geometry*Operator	6	68.3	11.378	3.26	0.003
Point(Geometry)	32	549.4	17.168	4.91	0.000
Rotation*Operator	2	100.9	53.508	15.31	0.000
Tool-device distance*Operator	2	211.9	105.952	30.31	0.000
Geometry*Rotation*Tool-device distance	6	313.5	104.513	29.51	0.000
Geometry*Rotation*Operator	6	60.6	10.100	2.93	0.008
Geometry*Tool-device distance*Operator	2	10.8	5.412	1.57	0.001
Rotation*Tool-device distance*Operator	2	46.3	23.162	6.36	0.041
Rotation*Tool-device distance*Point(Geometry)	2	115.66	5.175	1.48	0.041
Error	2055	7183.4	3.496		
Lack-of-Fit	327	750.7	2.296	0.62	1.000
Pure Error	1728	6432.7	3.723		
Total	2159	12433.6			

Data availability

The raw data and processing code from this study are available and can be accessed at the following link. The data and codes are provided for review and reuse under the terms of the CC BY 4.0 license.

References

Alves, J.B., Marques, B., Ferreira, C., Dias, P., Santos, B.S., 2022. Comparing augmented reality visualization methods for assembly procedures. Virtual Real. 1–14. http://dx.doi.org/10.1007/s10055-021-00557-8.

Amza, C.G., Zapciu, A., Teodorescu, O., 2018. Augmented reality application for industrial non-destructive inspection training. In: AIP Conference Proceedings. Vol. 1932, AIP Publishing, http://dx.doi.org/10.1063/1.5024152.

2010. ASTM international, west conshohocken, PA, ASTM F2554-10 standard practice for measurement of positional accuracy of computer assisted surgical systems. http://dx.doi.org/10.1520/F2554-10, Available at: https://www.astm.org/f2554-10.html.

Atuegwu, N., Galloway, R., 2008. Volumetric characterization of the aurora magnetic tracker system for image-guided transorbital endoscopic procedures. Phys. Med. Biol. 53 (16), 4355. http://dx.doi.org/10.1088/0031-9155/53/16/009.

Bailey, T., Durrant-Whyte, H., 2006. Simultaneous localization and mapping (slam): Part ii. IEEE Robot. Autom. Mag. 13 (3), 108–117. http://dx.doi.org/10.1109/MRA.2006.1678144.

Bi, Z., Wang, L., 2010. Advances in 3d data acquisition and processing for industrial applications. Robot. Comput.-Integr. Manuf. 26 (5), 403–413. http://dx.doi.org/10.1016/j.rcim.2010.03.003.

Birkfellner, W., Watzinger, F., Wanschitz, F., Enislidis, G., Kollmann, C., Rafolt, D., Nowotny, R., Ewers, R., Bergmann, H., 1998. Systematic distortions in magnetic position digitizers. Med. Phys. 25 (11), 2242–2248. http://dx.doi.org/10.1118/1.598425.

Birlo, M., Edwards, P.E., Clarkson, M., Stoyanov, D., 2022. Utility of optical see-through head mounted displays in augmented reality-assisted surgery: A systematic review. Med. Image Anal. 77, 102361. http://dx.doi.org/10.1016/j.media.2022.102361.

Bräuer, T., Andreeva, T., Endler, M., Kisslinger, J., 2003. Accuracy requirements for the fabrication and assembly of wendelstein 7-x. In: 14th International Stellarator Workshop.

Cannavò, A., D'Alessandro, A., Maglione, D., Marullo, G., Zhang, C., Lamberti, F., 2020. Automatic generation of affective 3d virtual environments from 2d images. pp. 113–124. http://dx.doi.org/10.5220/0008951301130124.

Carfagni, M., Furferi, R., Governi, L., Servi, M., Uccheddu, F., Volpe, Y., 2017. On the performance of the intel SR300 depth camera: metrological and critical characterization. IEEE Sensors J. 17 (14), 4508–4519. http://dx.doi.org/10.1109/JSEN.2017.2703829.

Cartiaux, O., Docquier, P.-L., Paul, L., Francq, B.G., Cornu, O.H., Delloye, C., Raucent, B., Dehez, B., Banse, X., 2008. Surgical inaccuracy of tumor resection and reconstruction within the pelvis: an experimental study. Acta Orthop. 79 (5), 695–702. http://dx.doi.org/10.1080/17453670810016731.

Castillo, C.D., Jacobs, D.W., 2009. Using stereo matching with general epipolar geometry for 2d face recognition across pose. IEEE Trans. Pattern Anal. Mach. Intell. 31 (12), 2298–2304. http://dx.doi.org/10.1109/TPAMI.2009.123.

Chandel, K., Åhlén, J., Seipel, S., 2022. Augmented reality and indoor positioning in context of smart industry: A review. Manag. Prod. Eng. Rev. 13 (4), 72–87. http://dx.doi.org/10.24425/mper.2022.142396.

Chen, E.C., Lasso, A., Fichtinger, G., 2020. External tracking devices and tracked tool calibration. In: HandBook of Medical Image Computing and Computer Assisted Intervention. Elsevier, pp. 777–794. http://dx.doi.org/10.1016/B978-0-12-816176-0.00036-3.

Day, J.S., Dumas, G.A., Murdoch, D.J., 1998. Evaluation of a long-range transmitter for use with a magnetic tracking device in motion analysis. J. Biomech. 31 (10), 957–961. http://dx.doi.org/10.1016/s0021-9290(98)00089-x.

De Pace, F., Manuri, F., Sanna, A., 2018. Augmented reality in industry 4.0. Am. J. Comput. Sci. Inf. Technol 6 (1), 17. http://dx.doi.org/10.21767/2349-3917.100017.

Dengel, A., Devillers, L., Schaal, L.M., 2021. Augmented human and human-machine co-evolution: Efficiency and ethics. Reflections Artif. Intell. Humanit. 203–227. http://dx.doi.org/10.1007/978-3-030-69128-8_13.

Durrant-Whyte, H., Bailey, T., 2006. Simultaneous localization and mapping: part i. IEEE Robot. Autom. Mag. 13 (2), 99–110. http://dx.doi.org/10.1109/MRA.2006.1638022.

Eck, U., Pankratz, F., Sandor, C., Klinker, G., Laga, H., 2015. Precise haptic device co-location for visuo-haptic augmented reality. IEEE Trans. Vis. Comput. Graphics 21 (12), 1427–1441. http://dx.doi.org/10.1109/TVCG.2015.2480087.

Feiner, S.K., 2002. Augmented reality: A new way of seeing. Sci. Am. 286 (4), 48–55. http://dx.doi.org/10.1038/scientificamerican0402-48.

Frantz, D.D., Wiles, A.D., Leis, S., Kirsch, S., 2003. Accuracy assessment protocols for electromagnetic tracking systems. Phys. Med. Biol. 48 (14), 2241. http://dx.doi.org/10.1088/0031-9155/48/14/314.

Franz, A.M., Haidegger, T., Birkfellner, W., Cleary, K., Peters, T.M., Maier-Hein, L., 2014. Electromagnetic tracking in medicine—a review of technology, validation, and applications. IEEE Trans. Med. Imaging 33 (8), 1702–1725. http://dx.doi.org/10.1109/TMI.2014.2321777.

Galetto, M., Genta, G., Maculotti, G., Verna, E., 2020. Defect probability estimation for hardness-optimised parts by selective laser melting. Int. J. Precis. Eng. Manuf. 21, 1739–1753. http://dx.doi.org/10.1007/s12541-020-00381-1.

Galetto, M., Verna, E., Genta, G., 2021. Effect of process parameters on parts quality and process efficiency of fused deposition modeling. Comput. Ind. Eng. 156, 107238. http://dx.doi.org/10.1016/j.cie.2021.107238.

Gao, Y., Lin, L., Chai, G., Xie, L., 2019. A feasibility study of a new method to enhance the augmented reality navigation effect in mandibular angle split osteotomy. J. Cranio-Maxillofacial Surg. 47 (8), 1242–1248. http://dx.doi.org/10.1016/j.jcms.2019.04.005.

Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F.J., Marín-Jiménez, M.J., 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognit. 47 (6), 2280–2292. http://dx.doi.org/10.1016/j.patcog.2014.01.005.

Genta, G., Maculotti, G., 2021. Uncertainty evaluation of small wear measurements on complex technological surfaces by machine vision-aided topographical methods. CIRP Ann 70 (1), 451–454. http://dx.doi.org/10.1016/j.cirp.2021.04.057.

2021. ISO 10360-13:2021 Geometrical Product Specifications (Gps) – Acceptance and Reverification Tests for Coordinate Measuring Systems (CMS) – Part 13: Optical 3D CMS, StandArd. International Organization for Standardization, Geneva, CH.

Gergel, I., Gaa, J., Müller, M., Meinzer, H.-P., Wegner, I., 2012. A novel fully automatic system for the evaluation of electromagnetic tracker. In: Medical Imaging 2012: Image-Guided Procedures, Robotic Interventions, and Modeling. Vol. 8316, SPIE, pp. 95–104. http://dx.doi.org/10.1117/12.911387.

Groves, L.A., Carnahan, P., Allen, D.R., Adam, R., Peters, T.M., Chen, E.C., 2019. Accuracy assessment for the co-registration between optical and vive head-mounted display tracking. Int. J. Comput. Assist. Radiol. Surg. 14, 1207–1215. http://dx.doi.org/10.1007/s11548-019-01992-4.

Haidegger, T., Fenyvesi, G., Sirokai, B., Kelemen, M., Nagy, M., Takács, B., Kovács, L., Benyó, B., Benyó, Z., 2011. Towards unified electromagnetic tracking system assessment-static errors. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 1905–1908. http://dx.doi.org/10.1109/IEMBS.2011.6090539.

Herregodts, S., Verhaeghe, M., De Coninck, B., Forward, M., Verstraete, M.A., Victor, J., De Baets, P., 2021. An improved method for assessing the technical accuracy of optical tracking systems for orthopaedic surgical navigation. Int. J. Med. Robot. Comput. Assist. Surg. 17 (4), e2285. http://dx.doi.org/10.1002/rcs.2285.

Ho, P.T., Albajez, J.A., Santolaria, J., Yagüe-Fabra, J.A., 2022. Study of augmented reality based manufacturing for further integration of quality control 4.0: A systematic literature review. Appl. Sci. 12 (4), 1961. http://dx.doi.org/10.3390/app12041961.

Huang, B.-C., Hsu, J., Chu, E.T.-H., Wu, H.-M., 2020. Arbin: Augmented reality based indoor navigation system. Sensors 20 (20), 5890. http://dx.doi.org/10.3390/s20205890.

Huey, L.C., Sebastian, P., Drieberg, M., 2011. Augmented reality based indoor positioning navigation tool. In: 2011 IEEE Conference on Open Systems. IEEE, pp. 256–260. http://dx.doi.org/10.1109/ICOS.2011.6079276.

Hummel, J.B., Bax, M.R., Figl, M.L., Kang, Y., Maurer, Jr., C., Birkfellner, W.W., Bergmann, H., Shahidi, R., 2005. Design and application of an assessment protocol for electromagnetic tracking systems. Med. Phys. 32 (7Part1), 2371–2379. http://dx.doi.org/10.1118/1.1944327.

Illinois Motion Group, 2024. Coordinate transformations. https://motion.cs.illinoisedu/RoboticSystems/CoordinateTransformations.html#Working-with-coordinates (Accessed 25 2024).

Innocente, C., Nonis, F., Lo Faro, A., Ruggieri, R., Ulrich, L., Vezzetti, E., 2024. A metaverse platform for preserving and promoting intangible cultural heritage. Appl. Sci. 14 (8), 3426. http://dx.doi.org/10.3390/app14083426.

Innocente, C., Piazzolla, P., Ulrich, L., Moos, S., Tornincasa, S., Vezzetti, E., 2022. Mixed reality-based support for total hip arthroplasty assessment. In: International Joint Conference on Mechanics, Design Engineering & Advanced Manufacturing. Springer, pp. 159–169. http://dx.doi.org/10.1007/978-3-031-15928-2_14.

Intel Corporation, 2019. RealSense SR30x product datasheet, datasheet. Available at: https://www.intelrealsense.com/wp-content/uploads/2019/07/RealSense_SR30x_Product_Datasheet_Rev_002.pdf.

2020. ISO 10360-1:2020 Geometrical Product Specifications (Gps) – Acceptance and Reverification Tests for Coordinate Measuring Machines (CMM) – Part 1: Vocabulary, StandArd. International Organization for Standardization, Geneva, CH.

2011. ISO 14253-2:2011 Geometrical Product Specifications (Gps) – Inspection By Measurement of Workpieces and Measuring Equipment – Part 2: Guidance for the Estimation of Uncertainty in GPS Measurement, in Calibration of Measuring Equipment and in Product Verification, StandArd. International Organization for Standardization, Geneva, CH.

2017. ISO 14253-1:2017 Geometrical Product Specifications (gps) – Inspection by measurement of workpieces and measuring equipment – Part 1: Decision rules for verifying conformity or nonconformity with specifications, Standard. International Organization for Standardization, Geneva, CH.

2021. ISO/ASTM 52900:2021 Additive Manufacturing – General Principles – Terminology, StandArd. International Organization for Standardization, Geneva, CH.

2008. JCGM 200:2008 International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM), StandArd. Joint Committee for Guides in Metrology - International Organization for Standardization, Geneva, CH.

2008. JCGM 100:2008 Evaluation of Measurement Data – Guide To the Expression of Uncertainty in Measurement (GUM), StandArd. Geneva, CH.

Jurado-Rodriguez, D., Muñoz-Salinas, R., Garrido-Jurado, S., Medina-Carnicer, R., 2023. Planar fiducial markers: a comparative study. Virtual Real. 27 (3), 1733–1749. http://dx.doi.org/10.1007/s10055-023-00772-5.

Kavitha, A., Kumar, S.P., Elona, J.J., Sudhir, G., 2025. Surgical tool tracking using augmented reality and infrared markers – a comparative study. Biomed. Sci. Instrum. 61 (1), 91–97. http://dx.doi.org/10.34107/UKKK6693.091.

Kholkhujaev, J., Maculotti, G., Genta, G., Galetto, M., Inoyatkhodjaev, J., 2022. Non-contact articulated robot-integrated gap and flushness measurement system for automobile assembly. IEEE Access 10, 86528–86541. http://dx.doi.org/10.1109/ACCESS.2022.3199066.

Koivukangas, T., Katisko, J.P., Koivukangas, J.P., 2013. Technical accuracy of optical and the electromagnetic tracking systems. SpringerPlus 2, 1–7. http://dx.doi.org/10.1186/2193-1801-2-90.

Korak, G., Kucera, G., 2015. Optical tracking system. Int. J. Electron. Telecommun. 61 (2), 165–170. http://dx.doi.org/10.1515/eletel-2015-0021.

Kügler, D., Krumb, H., Bredemann, J., Stenin, I., Kristin, J., Klenzner, T., Schipper, J., Schmitt, R., Sakas, G., Mukhopadhyay, A., 2019. High-precision evaluation of electromagnetic tracking. Int. J. Comput. Assist. Radiol. Surg. 14, 1127–1135. http://dx.doi.org/10.1007/s11548-019-01959-5.

Kunz, C., Maurer, P., Kees, F., Henrich, P., Marzi, C., Hlaváč, M., Schneider, M., Mathis-Ullrich, F., 2020. Infrared marker tracking with the hololens for neurosurgical interventions. In: Gruyter, De (Ed.), Current Directions in Biomedical Engineering. Vol. 6, 20200027. http://dx.doi.org/10.1515/cdbme-2020-0027.

Lee, S.-J., Yoo, J.-Y., Yoo, S.-K., Ha, R., Lee, D.-H., Kim, S.-T., Yi, W.-J., 2021. Image-guided endoscopic sinus surgery with 3d volumetric visualization of the nasal cavity and paranasal sinuses: A clinical comparative study. Appl. Sci. 11 (8), 3675. http://dx.doi.org/10.3390/app11083675.

Liebmann, F., Roner, S., von Atzigen, M., Scaramuzza, D., Sutter, R., Snedeker, J., Farshad, M., Fürnstahl, P., 2019. Pedicle screw navigation using surface digitization on the microsoft hololens. Int. J. Comput. Assist. Radiol. Surg. 14, 1157–1165. http://dx.doi.org/10.1007/s11548-019-01973-7.

Longuet-Higgins, H.C., 1981. A computer algorithm for reconstructing a scene from two projections. Nature 293 (5828), 133–135. http://dx.doi.org/10.1038/293133a0.

Lugez, E., Sadjadi, H., Pichora, D.R., Ellis, R.E., Akl, S.G., Fichtinger, G., 2015. Electromagnetic tracking in surgical and interventional environments: usability study. Int. J. Comput. Assist. Radiol. Surg. 10, 253–262. http://dx.doi.org/10.1007/s11548-014-1110-0.

Maculotti, G., Genta, G., Galetto, M., 2024. Optimisation of laser welding of deep drawing steel for automotive applications by machine learning: A comparison of different techniques. Qual. Reliab. Eng. Int. 40 (1), 202–219. http://dx.doi.org/10.1002/qre.3377.

Maculotti, G., Ulrich, L., Olivetti, E.C., Genta, G., Marcolin, F., Vezzetti, E., Galetto, M., 2022. A methodology for task-specific metrological characterization of low-cost 3d camera for face analysis. Measurement 200, 111643. http://dx.doi.org/10.1016/j.measurement.2022.111643.

Maier-Hein, L., Franz, A., Meinzer, H.-P., Wolf, I., 2008. Comparative assessment of optical tracking systems for soft tissue navigation with fiducial needles. In: Medical Imaging 2008: Visualization, Image-Guided Procedures, and Modeling. Vol. 6918, SPIE, pp. 664–672. http://dx.doi.org/10.1117/12.769181.

Martin-Gomez, A., Li, H., Song, T., Yang, S., Wang, G., Ding, H., Navab, N., Zhao, Z., Armand, M., 2023. Sttar: surgical tool tracking using off-the-shelf augmented reality head-mounted displays. IEEE Trans. Vis. Comput. Graphics http://dx.doi.org/10.48550/arXiv.2208.08880.

Masood, T., Egger, J., 2020. Adopting augmented reality in the age of industrial digitalisation. Comput. Ind. 115, 103112. http://dx.doi.org/10.1016/j.compind.2019.07.002.

MathWorks, 2024a. Camera calibration patterns. https://it.mathworkscom/help/vision/ug/calibration-patterns.html (Accessed 20 2024).

MathWorks, 2024b. Using the stereo camera calibrator app. https://it.mathworks.com/help/vision/ug/using-the-stereo-camera-calibrator-app.html#buejbxy-13 (Accessed 18 2024).

Maurer, C.R., Fitzpatrick, J.M., Wang, M.Y., Galloway, R.L., Maciunas, R.J., Allen, G.S., 1997. Registration of head volume images using implantable fiducial markers. IEEE Trans. Med. Imaging 16 (4), 447–462. http://dx.doi.org/10.1109/42.611354.

Milgram, P., Colquhoun, H., et al., 1999. A taxonomy of real and virtual world display integration. Mix. Real.: Merging Real Virtual Worlds 1 (1999), 1–26. http://dx.doi.org/10.1007/978-3-642-87512-0_1.

Milgram, P., Takemura, H., Utsumi, A., Kishino, F., 1995. Augmented reality: A class of displays on the reality-virtuality continuum. In: Telemanipulator and Telepresence Technologies. Vol. 2351, Spie, pp. 282–292. http://dx.doi.org/10.1117/12.197321.

Minetola, P., Calignano, F., Galati, M., 2020. Comparing geometric tolerance capabilities of additive manufacturing systems for polymers. Addit. Manuf. 32, 101103. http://dx.doi.org/10.1016/j.addma.2020.101103.

Montgomery, D.C., 2017. Design and Analysis of Experiments. John wiley & sons.

Murphy, M.J., Eidens, R., Vertatschitsch, E., Wright, J.N., 2008. The effect of transponder motion on the accuracy of the calypso electromagnetic localization system. Int. J. Radiat. Oncology* Biology* Phys. 72 (1), 295–299. http://dx.doi.org/10.1016/j.ijrobp.2008.05.036.

Nakano, G., 2016. A versatile approach for solving pnp, pnpf, and pnpfr problems. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the NetherlandS, October (2016) 11-14, Proceedings, Part III 14. Springer, pp. 338–352. http://dx.doi.org/10.1007/978-3-319-46487-9_21.

Narasimhan, S., Turkcan, M.K., Ballo, M., Choksi, S., Filicori, F., Kostic, Z., 2025. Monocular 3D tooltip tracking in robotic surgery—Building a multi-stage pipeline. Electron. 14 (10), 2075. http://dx.doi.org/10.3390/electronics14102075.

Northern Digital Inc, 2020. Aurora electromagnetic tracking system, datasheet. Available at: https://ndicorpstg.wpengine.com/wp-content/uploads/2020/09/10006091-Aurora-Tear-Sheet.pdf.

OpenCV Community, 2024. Charuco board detection. https://docs.opencv.org/3.4/df/d4a/tutorial_charuco_detection.html (Accessed 21 2024).

Parker, J.R., 2010. Algorithms for Image Processing and Computer Vision. John Wiley & Sons.

Pietruski, P., Majak, M., Światek-Najwer, E., Zuk, M., Popek, M., Mazurek, M., Świecka, M., Jaworowski, J., 2019. Supporting mandibular resection with intraoperative navigation utilizing augmented reality technology–a proof of concept study. J. Cranio-Maxillofacial Surg. 47 (6), 854–859. http://dx.doi.org/10.1016/j.jcms.2019.03.004.

Pirmagomedov, R., Koucheryavy, Y., 2021. Iot technologies for augmented human: A survey. Internet Things 14, 100120. http://dx.doi.org/10.1016/j.iot.2019.100120.

Radkowski, R., Kanunganti, S., 2018. Augmented reality system calibration for assembly support with the microsoft hololens. In: International Manufacturing Science and Engineering Conference. Vol. 51371, American Society of Mechanical Engineers, V003T02A021. http://dx.doi.org/10.1115/MSEC2018-6660.

Raisamo, R., Rakkolainen, I., Majaranta, P., Salminen, K., Rantala, J., Farooq, A., 2019. Human augmentation: Past, present and future. Int. J. Hum.-Comput. Stud. 131, 131–143. http://dx.doi.org/10.1016/j.ijhcs.2019.05.008.

Ramírez, H., Mendoza, E., Mendoza, M., González, E., 2015. Application of augmented reality in statistical process control, to increment the productivity in manufacture. Procedia Comput. Sci. 75, 213–220. http://dx.doi.org/10.1016/j.procs.2015.12.240.

Rampersaud, Y.R., Simon, D.A., Foley, K.T., 2001. Accuracy requirements for image-guided spinal pedicle screw placement. Spine 26 (4), 352–359. http://dx.doi.org/10.1097/00007632-200102150-00010.

Rauschnabel, P.A., Babin, B.J., tom Dieck, M.C., Krey, N., Jung, T., 2022. What is augmented reality marketing? its definition, complexity, and future. J. Bus. Res. 142, 1140–1150. http://dx.doi.org/10.1016/j.jbusres.2021.12.084.

Reichl, T., Gardiazabal, J., Navab, N., 2013. Electromagnetic servoing—a new tracking paradigm. IEEE Trans. Med. Imaging 32 (8), 1526–1535. http://dx.doi.org/10.1109/TMI.2013.2259636.

Rudolph, T., Ebert, L., Kowal, J., 2010. Comparison of three optical tracking systems in a complex navigation scenario. Comput. Aided Surg. 15 (4–6), 104–109. http://dx.doi.org/10.3109/10929088.2010.530793.

Santanam, L., Noel, C., Willoughby, T.R., Esthappan, J., Mutic, S., Klein, E.E., Low, D.A., Parikh, P.J., 2009. Quality assurance for clinical implementation of an electromagnetic tracking system. Med. Phys. 36 (8), 3477–3486. http://dx.doi.org/10.1118/1.3158812.

Scholz, J., Smith, A.N., 2016. Augmented reality: Designing immersive experiences that maximize consumer engagement. Bus. Horiz. 59 (2), 149–161. http://dx.doi.org/10.1016/j.bushor.2015.10.003.

Settimi, A., Gamerro, J., Weinand, Y., 2022. Augmented-reality-assisted timber drilling with smart retrofitted tools. Autom. Constr. 139, 104272. http://dx.doi.org/10.1016/j.autcon.2022.104272.

Shen, E., Shechter, G., Kruecker, J., Stanton, D., 2007. Quantification of ac electromagnetic tracking system accuracy in a ct scanner environment. In: Medical Imaging 2007: Visualization and Image-Guided Procedures. Vol. 6509, SPIE, pp. 203–212. http://dx.doi.org/10.1117/12.710836.

Singh, S., Sharma, Y., Liaqat, A., Kalawsky, R.S., 2025. Evaluation of XR device's real-world tracking accuracy and depth perception from an industrial point of view. Virtual Real. 29 (3), 1–10. http://dx.doi.org/10.1007/s10055-025-01192-3.

Smith, R., Schwiegerling, J., 2023. Head mounted display based augmented reality device for medical applications. In: ODS 2023: Industrial Optical Devices and Systems. Vol. 12684, SPIE, pp. 102–106. http://dx.doi.org/10.1117/12.2677835.

Sorriento, A., Porfido, M.B., Mazzoleni, S., Calvosa, G., Tenucci, M., Ciuti, G., Dario, P., 2019. Optical and electromagnetic tracking systems for biomedical applications: A critical review on potentialities and limitations. IEEE Rev. Biomed. Eng. 13, 212–232. http://dx.doi.org/10.1109/RBME.2019.2939091.

Stradiotti, S., Emiliani, N., Marcelli, E., Cercenelli, L., 2024. Understanding how different visual aids for augmented reality influence tool-patient alignment in surgical tasks: A preliminary study. In: VISIGRAPP (1): GRAPP, HUCAPP, IVAPP. pp. 616–622. http://dx.doi.org/10.5220/0012611800003660.

Teatini, A., Pérez de Frutos, J., Eigl, B., Pelanis, E., Aghayan, D.L., Lai, M., Kumar, R.P., Palomar, R., Edwin, B., Elle, O.J., 2021. Influence of sampling accuracy on augmented reality for laparoscopic image-guided surgery. Minim. Invasive Ther. & Allied Technol. 30 (4), 229–238. http://dx.doi.org/10.1080/13645706.2020.1727524.

Ulrich, L., Nonis, F., Vezzetti, E., Moos, S., Caruso, G., Shi, Y., Marcolin, F., 2021. Can adas distract driver's attention? an rgb-d camera and deep learning-based analysis. Appl. Sci. 11 (24), 11587. http://dx.doi.org/10.3390/app112411587.

Ulrich, L., Salerno, F., Moos, S., Vezzetti, E., 2024. How to exploit augmented reality (ar) technology in patient customized surgical tools: a focus on osteotomies. Multimedia Tools Appl. 1–32. http://dx.doi.org/10.1007/s11042-023-18058-y.

Vávra, P., Roman, J., Zonča, P., Ihnát, P., Němec, M., Kumar, J., Habib, N., El-Gendi, A., 2017. Recent development of augmented reality in surgery: a review. J. Heal. Eng. 2017 (1), 4574172. http://dx.doi.org/10.1155/2017/4574172.

Wang, J., Meng, M.Q.-H., Ren, H., 2015. Towards occlusion-free surgical instrument tracking: A modular monocular approach and an agile calibration method. IEEE Trans. Autom. Sci. Eng. 12 (2), 588–595. http://dx.doi.org/10.1109/TASE.2015.2388537.

Wang, J., Zhang, T., Zhang, Z., Meng, M.Q.-H., Song, S., 2023. Tracking-by-registration: A robust approach for optical tracking system in surgical navigation. IEEE Trans. Instrum. Meas. http://dx.doi.org/10.1109/TIM.2023.3312484.

West, J.B., Khadem, R., Maurer Jr., C.R., 2003. Nonlinear averaging method to optimize the accuracy of optical tracking. In: Medical Imaging 2003: Visualization, Image-Guided Procedures, and Display. Vol. 5029, SPIE, pp. 311–318. http://dx.doi.org/10.1117/12.479697.

West, J.B., Maurer, C.R., 2004. Designing optically tracked instruments for image-guided surgery. IEEE Trans. Med. Imaging 23 (5), 533–545. http://dx.doi.org/10.1109/TMI.2004.825614.

Whitehead, A., Roth, G., 2004. Estimating intrinsic camera parameters from the fundamental matrix using an evolutionary approach. EURASIP J. Adv. Signal Process. 2004, 1–12. http://dx.doi.org/10.1155/S1110865704401024.

Wiles, A.D., Thompson, D.G., Frantz, D.D., 2004. Accuracy assessment and interpretation for optical tracking systems. In: Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display. Vol. 5367, SPIE, pp. 421–432. http://dx.doi.org/10.1117/12.536128.

Wilson, E., Yaniv, Z., Zhang, H., Nafis, C., Shen, E., Shechter, G., Wiles, A.D., Peters, T., Lindisch, D., Cleary, K., 2007. A hardware and software protocol for the evaluation of electromagnetic tracker accuracy in the clinical environment: a multi-center study. In: Medical Imaging 2007: Visualization and Image-Guided Procedures. Vol. 6509, SPIE, pp. 960–970. http://dx.doi.org/10.1117/12.712701.

Yang, R., Wang, Z., Liu, S., Wu, X., 2012. Design of an accurate near infrared optical tracking system in surgical navigation. J. Lightwave Technol. 31 (2), 223–231. http://dx.doi.org/10.1109/JLT.2012.2227943.

Yaniv, Z., 2015. Which pivot calibration? In: Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling. Vol. 9415, SPIE, pp. 542–550. http://dx.doi.org/10.1117/12.2081348.

Yaniv, Z., Wilson, E., Lindisch, D., Cleary, K., 2009. Electromagnetic tracking in the clinical environment. Med. Phys. 36 (3), 876–892. http://dx.doi.org/10.1118/1.3075829.

Yenduri, G., Maddikunta, P.K.R., Gadekallu, T.R., Jhaveri, R.H., Bandi, A., Chen, J., Wang, W., Shirawalmath, A.A., Ravishankar, R., Wang, W., et al., 2024. Spatial computing: Concept, applications, challenges and future directions. http://dx.doi.org/10.48550/arXiv.2402.07912, arXiv preprint arXiv:2402.07912.

Yoo, J., Schafer, S., Uneri, A., Otake, Y., Khanna, A.J., Siewerdsen, J.H., 2013. An electromagnetic tracker-in-table configuration for x-ray fluoroscopy and cone-beam ct-guided surgery. Int. J. Comput. Assist. Radiol. Surg. 8, 1–13. http://dx.doi.org/10.1007/s11548-012-0744-z.

Zhang, Z., 2000. A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. 22 (11), 1330–1334. http://dx.doi.org/10.1109/34.888718.

Zhou, C., Zhu, M., Shi, Y., Lin, L., Chai, G., Zhang, Y., Xie, L., 2017. Robot-assisted surgery for mandibular angle split osteotomy using augmented reality: preliminary results on clinical animal experiment. Aesthetic Plast. Surg. 41, 1228–1236. http://dx.doi.org/10.1007/s00266-017-0900-5.

Zhu, Z., Liu, C., Xu, X., 2019. Visualisation of the digital twin data in manufacturing by using augmented reality. Procedia Cirp 81, 898–903. http://dx.doi.org/10.1016/j.procir.2019.03.223.

Zollmann, S., Langlotz, T., Grasset, R., Lo, W.H., Mori, S., Regenbrecht, H., 2020. Visualization techniques in augmented reality: A taxonomy, methods and patterns. IEEE Trans. Vis. Comput. Graphics 27 (9), 3808–3825. http://dx.doi.org/10.1109/TVCG.2020.2986247.