

Robotic Characterization of Markerless Hand-Tracking on Meta Quest Pro and Quest 3 Virtual Reality Headsets

Eric Godden , William Steedman , and Matthew K.X.J. Pan 

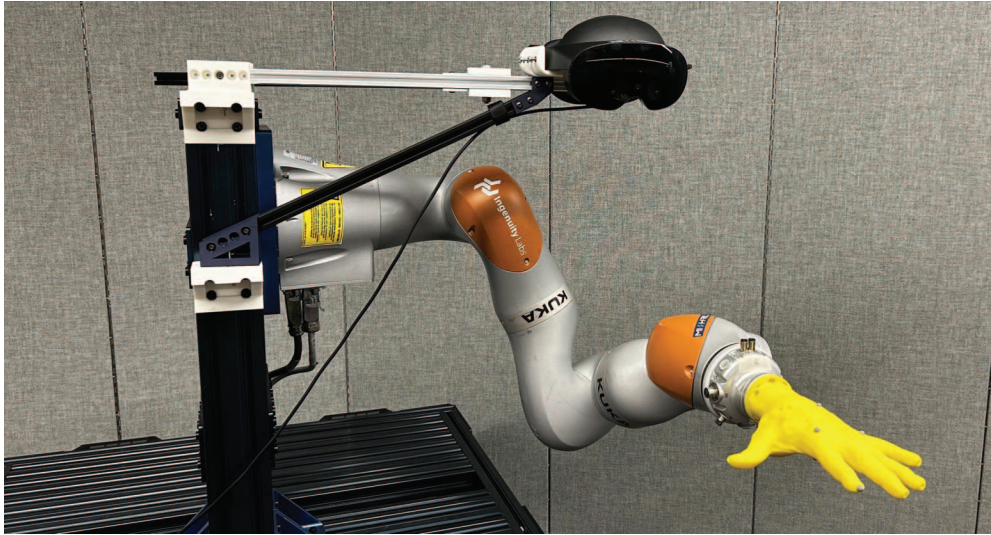


Fig. 1: Fully integrated markerless hand-tracking characterization system which combines the VR HMD (Meta Quest Pro shown), KUKA LBR iiwa 7 R800 robotic manipulator, and custom hand model end effector.

Abstract—Markerless hand-tracking has become increasingly common on commercially available virtual and mixed reality headsets to improve the naturalness of interaction and immersivity of virtual environments. However, there has been limited examination of the performance of markerless hand-tracking on commercial head-mounted displays. Here, we propose an evaluation methodology that leverages a robotic manipulator to measure the positional accuracy, jitter, and latency of such systems and provides a standardized characterization framework of markerless hand-tracking. We apply this methodology to evaluate the hand-tracking performance of two recent mixed reality devices from Meta: the Quest Pro and Quest 3. Results demonstrate the influence of proximity to the headset, rotation of hand, and joint selected as the tracking feature on hand-tracking performance. We found that hand-tracking error and jitter were lowest for both headsets in conditions where the knuckle was the tracking point compared to the fingertip. Regarding positional accuracy, in best-performing conditions, the Quest Pro outperformed the Quest 3 with 1.22 cm of average error compared to 1.73 cm. The opposite result was true concerning jitter, with results of 1.77 cm and 1.11 cm for the Quest Pro and Quest 3, respectively. We found latency highly variable for the Quest Pro (15.8 - 229.2 ms) and Quest 3 (14.4 - 220.5 ms). This work provides a testing framework for highly systematic and repeatable performance measurements of markerless hand-tracking systems embedded in headsets.

Index Terms—Hand-tracking, robotics, tracking performance, characterization, virtual reality, head-mounted display.

1 INTRODUCTION

Markerless hand-tracking is becoming a common feature on virtual reality (VR) head-mounted displays (HMDs) and is present on current generation Meta Quest and HTC Vive products. Hand-tracking, as opposed to controller-tracking, allows the user to interact with the virtual environment without needing handheld controllers. This ability produces more realistic and meaningful interactions with a virtual environment and can put the user in a more active role within their experience. VR interactions based on hand-tracking are becoming more commonplace for VR applications such as physical therapy [23] and robotic teleoperation [25].

In these and other applications, an understanding of hand-tracking performance is essential to users, practitioners, and developers to understand their limitations; hand-tracking is still a developing technology within the VR domain as limitations of tracking location and performance persist [4]. The technical difficulty of markerless hand-tracking over the traditional approach of having a user hold instrumented controllers necessitates computer vision methods for processing inputs, thus leading to lag and added noise that may hinder the overall performance of the system [32]. Without an understanding of hand-tracking performance, there can be a mismatch between the expected and actual accuracy of hand position data, which may be detrimental to the user experience. For example, hand-tracking fidelity is critical to virtual and mixed-reality experiences that merge real and virtual worlds: in [28], accurate hand-tracking (via motion capture) is needed to allow a user to catch a ball while immersed in VR. Also, in [19], markerless hand-tracking is used within a virtual experience to evaluate and classify tremors in the hands of patients with Parkinson's disease. Sufficient hand-tracking accuracy is required in evaluating tremors to ensure highly detailed data can be captured.

Here, we focus on this need to characterize the markerless hand-

Eric Godden, William Steedman, and Matthew K.X.J. Pan are with Queen's University. Email: {18eg16|21wks1|matthew.pan}@queensu.ca.

Received 18 September 2024; revised 13 January 2025; accepted 13 January 2025.
Date of publication 7 March 2025; date of current version 31 March 2025.
Digital Object Identifier no. 10.1109/TVCG.2025.3549182

tracking performance and present a process for doing so. In this work, we describe a robotic testing methodology (our testing setup is shown in Figure 1) to provide standardized testing of markerless hand-tracking performance. Using this process, we obtain measurements of positional accuracy, jitter, and latency for two commercially available VR HMDs: the Meta Quest Pro and Quest 3 HMDs. We posit that our HMD characterization will be a foundational step towards understanding, addressing, and improving markerless hand-tracking for VR and other systems, improving the immersivity of future HMDs.

2 BACKGROUND

2.1 Markerless Tracking

While it is clear hand-tracking can be beneficial to user experience in VR, markerless tracking is a difficult task that presents many additional challenges compared to traditional controller tracking. Controllers within VR systems are integrated with additional pieces of technology, such as inertial measurement units (IMU) and infrared lights, to enable sensor fusion and optical tracking to enhance performance [12]. Markerless tracking, on the other hand, inherently does not rely on any augmentations to the tracked object and must perform its tracking from an outside perspective. This technique often requires computationally expensive computer vision algorithms and demands highly detailed image input [31]. In addition to the use of more limited technology for tracking, the hand is a difficult object to track. The hand has more than 20 degrees of freedom, can experience rapid motions up to 5 m/s and 300 %/s, and experiences self-occlusions [8]. These factors often result in decreased performance compared to traditional controller tracking, which takes the form of lag within the user experience, and highlight the critical need for mitigation strategies in the event of tracking failure [9].

Common techniques of markerless hand-tracking use RGB-D [26] and infrared cameras [29] to obtain the information about the hand necessary for tracking. In recent years, learning techniques have leveraged neural networks to train algorithms, which have improved hand-tracking performance. One example of this includes the work by Han et al. [11], which serves as the default for hand-tracking on Meta Quest headsets. This approach uses four monochrome cameras to provide multiple image inputs at each tracking frame. An efficient convolutional neural network simultaneously localizes and classifies the hands within the images. A motion prediction step is also present in this algorithm to help minimize system latency. The final step predicts 21 key points across the hand to provide detailed information on the anatomical features of the hand.

2.2 Tracking Performance

Previous studies have evaluated the effects of latency within human-machine interfaces. This knowledge is foundational for performance analyses in VR. Lee et al. [17] explored the effect of delays in a networked human-machine system; results show an average discrimination threshold for visual delays to be 50 ms and a greater requirement of simultaneity when haptic information is involved. Jay et al. [15] also explored the effect of visual lag in a reciprocal tapping task, determining that the discernible threshold for visual lag is 94 ms. This result represents performance degradation with increased latency under a simple visual task, which may only become more severe and/or sensitive in the complex interactions experienced within VR applications. Significant correlations between latency and other factors have also been shown in VR, such as decreased bandwidth [20], motor performance [32], body ownership and inducing cybersickness [5] with increased visual delay. These effects cover many aspects of user experience in virtual environments, and the latency of commercial HMDs must be considered when developing virtual experiences to produce desired outcomes.

Prior work focusing on performance analyses of VR systems includes studies on room-scale tracking, controller, and hand-tracking. Room-scale tracking characterizations are generally performed with devices that use lighthouse tracking systems [13, 27], as performance can vary with proximity to base stations. Robotic manipulators have been used in VR device characterizations [7, 16] due to their ability to establish repeatable motion and the potential to establish ground-truth reference. Motion capture technology can also establish ground-truth

reference and is efficient for use in room-scale analyses [3]. Mizera et al. [24] evaluate two hand-tracking wearables and a vision-based system for use in teleoperation and virtual manipulation; the devices studied in this work have the drawbacks of requiring additional hardware to accompany a user's VR HMD.

As markerless hand-tracking technology is relatively new in commercial VR devices, there is limited prior work in this area. To our knowledge, Abdulkarim et al. [1] represent the only markerless hand-tracking characterization of a commercial HMD. In their work, participants were outfitted with reflective markers on their index finger to compare motion-capture data to the hand-tracking of the Meta Quest 2. The characterization consisted of a target-reaching and hand opening-and-closing task to produce results of delay (referred to as latency in our work), static positional error, path offset, and angular error of the index finger. Their work produced a single average delay result of 45 ms by determining the time difference in velocity peaks between data from the two measurement systems. Additionally, they found an average static positional error of 1.1 cm, and proximity to the HMD influenced the tracking error.

Our work builds upon the markerless hand-tracking methodology presented by Abdulkarim et al. by developing a more rigid set of tests that uses a robotic manipulator to characterize markerless hand-tracking on more recent HMDs.

3 OBJECTIVES

The objectives of this study are three-fold:

- Design a repeatable and extensible testing process that enables robust characterization of markerless hand-tracking performance in terms of latency, pose accuracy, and pose jitter.
- Characterize the performance of two current generation HMDs equipped with onboard markerless hand-tracking, the Meta Quest Pro and Meta Quest 3, using this testing process.
- Understand the limitations of markerless hand-tracking on these devices in terms of the observed metrics and discuss their impact on implementing applications that benefit from hand-tracking.

4 OVERVIEW OF FRAMEWORK FOR MARKERLESS HAND-TRACKING CHARACTERIZATION

We propose a comprehensive framework to systematically evaluate the performance of markerless hand-tracking systems in VR HMDs. By combining robotics, motion capture, and virtual environment rendering, the framework enables precise and repeatable assessments of key metrics such as positional accuracy, jitter, and latency. It addresses the need for standardized evaluation methodologies to ensure reliable comparisons across devices and conditions.

The framework integrates several key components, as shown in Figure 2. A robotic manipulator (we use a KUKA LBR iiwa 7 R800) performs repeatable and precise movements to simulate hand motions, while a 3D-printed hand model with retroreflective markers provides standardized and accurate ground-truth reference points. The Vicon motion capture system delivers sub-millimeter precision in tracking the hand model, enabling comparisons with the hand-tracking data from HMDs like the Meta Quest Pro and Quest 3. All subsystems, including the virtual environment rendered in Unity 3D, are synchronized through real-time communication protocols to ensure consistency.

The framework evaluates performance through three tests. The static performance test measures accuracy and jitter under stationary conditions by positioning the hand model at various points within the tracking volume. The dynamic performance test assesses latency during motion, calculating metrics like cold-start and constant-velocity latency by moving the hand along predefined trajectories. The motion-to-photon latency test measures the end-to-end delay from hand motion to visual representation on the HMD, offering insights into user-perceived latency.

This framework introduces several innovations, including the use of robotic precision for repeatability and the evaluation of multiple metrics for a holistic performance analysis. Its modular design allows for future

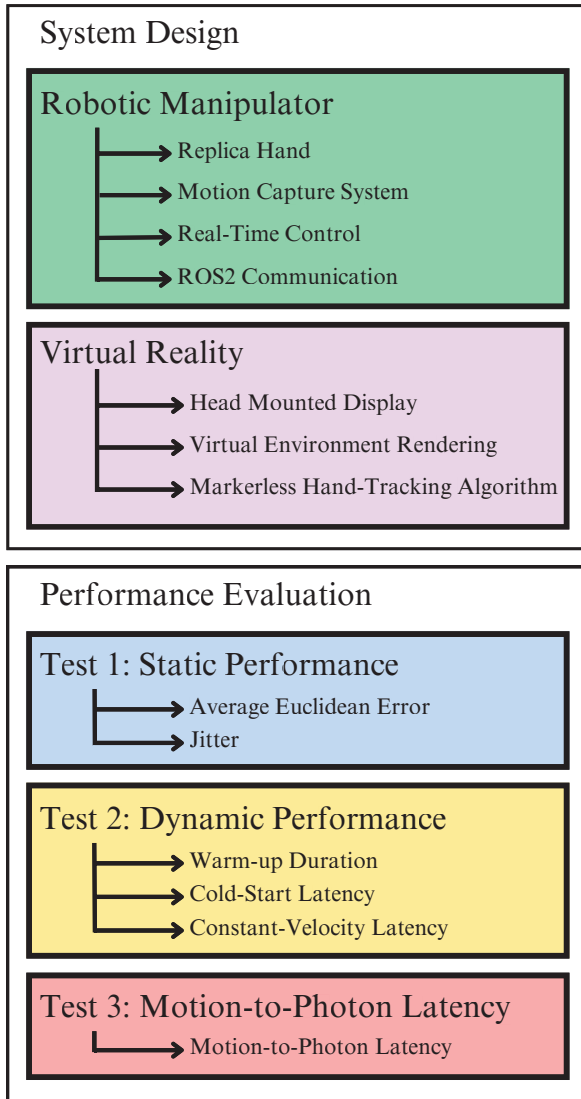


Fig. 2: Diagram of the markerless hand-tracking characterization framework developed in this work. The framework integrates robotic manipulator and virtual reality subsystems to assess tracking performance through three tests, with the corresponding results outlined for each.

adaptations, such as expanding the testing volume or incorporating additional metrics, such as complex gestures by way of an articulated hand model. By standardizing the characterization of markerless hand-tracking systems, this framework provides a foundation for developing more reliable and immersive VR technologies.

5 EXPERIMENTAL SETUP

The experimental setup we developed in this work for testing hand-tracking performance combines several subsystems: the VR HMDs, motion capture system, robot manipulator, and underlying communication system that allows data to be transferred between the aforementioned subsystems.

5.1 Virtual Reality Head Mounted Displays

The two HMDs evaluated in this study are the Meta Quest Pro¹ and the Meta Quest 3². The Meta Quest HMDs were selected due to their popularity, influence on the VR community, and their onboard

markerless hand-tracking capabilities. The Meta Quest Pro features the Qualcomm Snapdragon XR2+ Gen 1 processor, with 12 GB of RAM and five integrated cameras to facilitate tracking. The Meta Quest 3 features the Qualcomm Snapdragon XR2 Gen 2 processor with 8 GB of RAM and four integrated cameras with a depth sensor to facilitate tracking.

The virtual environment used for testing was developed and rendered using the Unity 3D game engine (2022.3.17f1) on a Windows 11 x64 laptop (AMD Ryzen 7 6800H 3.2GHz, 16GB RAM, NVidia Geforce RTX 3050 Ti). The virtual environment included a minimalistic plane of neutral colour to create a scene of low complexity to minimize the rendering workload. All testing was done with the HMDs tethered via a USB 3.0 wired connection to the laptop rather than running the HMDs in standalone mode to limit latency in data transmission and collection. Hand-tracking pose data was accessed through Unity's XR Interaction Toolkit³ and transmitted at a rate of approximately 90 Hz (transmission occurs within Unity's `MonoBehaviour.Update()` function that is variable and depends on system performance). 90 Hz was chosen as the refresh rate as it is the highest shared refresh rate available between the pair of headsets.

5.2 Robotic Manipulator

We used a robotic manipulator for our study to produce repeatable motions essential for comparison across trials and hardware. In our experimental setup, a KUKA LBR iiwa 7 R800 robot manipulator⁴, horizontally mounted on a pedestal, was used to mimic the kinematics of a human arm and precisely perform the stationary and dynamic positioning required for this characterization. The robot was position-controlled and received pose data in real-time with a loop rate of 1000 Hz using KUKA's proprietary Fast Robot Interface (FRI) [30].

The robot was controlled via a desktop machine (Intel Core i7-14700 4.2GHz, 32GB RAM) with Ubuntu 22.04 real-time kernel. Software for driving testing behaviours on the robot was developed upon the Robot Operating System 2 (ROS 2) Humble Hawksbill version, with the `lbr_fri_ros2_stack` being used to provide ROS 2 packages for FRI communication with the robot [14]. A custom mechanical support structure was designed to fix the headset location and reinforced to ensure vibrations through the structure were minimized.

5.3 Replica Hand

Since this work focuses on hand-tracking performance, a standardized hand was fabricated to provide consistency across all testing trials. An open-source hand model⁵, produced through an Eva optical scan, was used, which ensured high-accuracy and clear anatomical features which were required for key point matching on the virtual hand. This model was further modified to allow for 6.4 mm spherical reflective markers to be placed at each anatomical joint on the dorsal and palmar side of the hand, as well as at each fingertip. The tracking location of each joint was the midpoint between corresponding dorsal and palmar side markers. The model⁶ was 3D-printed using PLA material on an FDM printer, as shown in Figure 3. This technique provided high confidence in defining ground-truth locations in relation to the corresponding key points of the VR hand-tracking. A hand was fixed to the manipulator's end-effector and secured consistently for each trial. Pilot testing demonstrated no degradation in landmark identification or hand-tracking capabilities due to the added markers and filament colour. The combined system is shown in Figure 1.

5.4 Motion Capture System

We used a retroreflective marker-based motion-capture system, Vicon Vero (v2.2)⁷, to provide an efficient ground-truth reference for our VR

³<https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@3.0/manual/index.html>

⁴<https://www.kuka.com/en-ca/products/robotics-systems/industrial-robots/lbr-iiwa>

⁵<https://grabcad.com/library/anatomy-model-hand-1>

⁶<https://mithrilab.com/research/robotic-haptic-proxies-in-virtual-reality/robotic-characterization-of-markerless-hand-tracking/>

⁷<https://www.vicon.com/hardware/cameras/vero/>

¹<https://www.meta.com/ca/quest/quest-pro/>

²<https://www.meta.com/ca/quest/quest-3/>

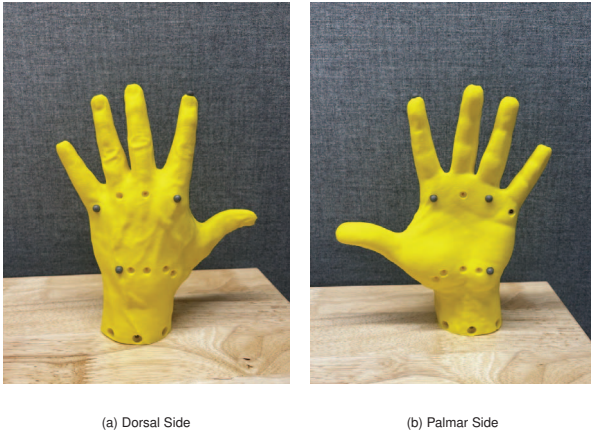


Fig. 3: Hand model with reflective markers at key anatomical features to define ground truth within the motion-capture system.

hand-tracking data to be analyzed. The system was used in tandem with the software Vicon Tracker (3.9.0.128020h)⁸ that was run on a Windows 11 x64 system (Intel Xeon W-2123 3.6GHz, 16GB RAM) to provide tracking of retroreflective marker constellations. This system produces sub-millimeter error and a low camera latency of 3.6 ms.

5.5 Communication

The synchronization of all subsystems was coordinated on the robot controller computer, facilitating the manipulator control. To synchronize efficiently, the VR and motion-capture systems delivered serialized data, using Google's Protocol Buffer Framework⁹, through high-speed UDP over ethernet. This minimized system latency and the influence of network traffic.

5.6 Registration Between Measurement Frames

To ensure proper comparison of data between the motion-capture and VR coordinate frames, a registration process was completed. This process involved taking measurements of the same anatomical feature on the hand over 30 unique points in both frames. A least squares approach was used to produce a homogeneous transformation matrix where the two sets of values were adjusted around their respective means and the singular value decomposition was taken.

6 METHODOLOGY

Our testing methodology examines markerless hand-tracking accuracy, jitter, and latency for both the Meta Quest Pro and Quest 3 through three tests: a static performance test, where hand-tracking performance is evaluated when the hand is not moving; a dynamic performance test, where hand-tracking performance is evaluated as the hand is moving; and, a hand-to-photon latency test that measures the delay between hand motion and when that motion is displayed on the HMD.

While other characterizations may only report single-value results for performance factors (e.g., [1]), this work aims to provide detailed results that contain multiple evaluations for both positional and latency performance. This will produce a more comprehensive understanding of hand-tracking performance that is currently unavailable.

6.1 Test 1: Static Performance

This test evaluates hand-tracking performance under static conditions, removing the influence of latency on measurement accuracy. This scenario represents the optimal tracking condition, as temporal factors are eliminated, and the hand's pose remains constant between frames. This consistency provides a stable perspective to the HMD cameras, simplifying the identification step in the tracking algorithm. The static

test condition is crucial for understanding the upper limit of tracking performance and is particularly relevant for precise interactions in virtual environments [22]. Abdulkarim et al. [1] highlighted the influence of hand proximity on tracking error. This motivates the static testing to be performed across a three-dimensional testing volume defined by the lateral and anterior proximity to the HMD. For each level of consistent vertical proximity, there are seven base positions, as shown in Figure 4. Table 1 details the lateral and anterior proximity to the HMD and the yaw rotation, as an offset from the pose with fingertips facing forward, for each position. Positions 1, 4, and 6 were adjusted through yaw rotations to recreate more natural hand positions that would be found using an anatomically accurate manipulator. The positions were chosen to evenly cover a quarter of a circular testing area with a radius of 0.4 m, centered on the first position.

Position	Lat. Offset (m)	Vert. Offset (m)	Yaw Rotation (°)
1	0.1	0.15	45
2	0.3	0.15	0
3	0.241	0.291	0
4	0.1	0.35	22.5
5	0.5	0.15	0
6	0.383	0.433	-15
7	0.1	0.55	0

Table 1: Static test positions for each consistent vertical level of three-dimensional testing volume defined by the lateral and anterior proximity to the HMD and yaw rotation.

Three vertical levels were defined to extend the testing volume into a third dimension. The first level, situated closest to the HMD, was positioned approximately 0.1m below eye level. The following two levels were positioned 0.2m and 0.4m below the first, creating a testing volume that is a quarter section of a 0.4m radius and 0.4m tall cylinder. Similar to the adjustment in yaw rotation of the hand across the seven positions, all of the positions on the first level were pitched upwards. The first three positions were pitched 45 degrees upward from horizontal, and the rest were pitched 22.5 degrees. This, again, ensures that the hand is positioned in an anatomically realistic position at elevated levels and provides greater line-of-sight visibility between the HMD and the tracked index finger features. As the manipulator and hand model were both in a left-hand configuration, only the region to the left of the HMD was tested. The testing region was defined solely to the left of the HMD as the majority of left-handed interactions within VR would occur in this region, giving it priority over extended workspaces that would rarely inhabit a left-handed interaction. We assume that performance is symmetric across the medial-lateral plane of the headset, allowing the results of this work to be extended to a right-handed model in its corresponding most relevant regions as well.

To further improve the robustness of this test, seven rotations were applied to each base position before moving to the next. As hands will rarely sit flat during interactions within a virtual environment, the applied rotations will ensure the results within this test consider the influence of rotation on hand-tracking performance. The rotations included (in this order denoted 1 through 7): an identity rotation (to keep hand at base position), ulnar and radial deviation of 15 degrees, flexion and extension of 45 degrees, and supination and pronation of 15 degrees. The flexion/extension conditions were given greater rotation as that was required to have a sufficient change in hand profile from the perspective of the HMD. With seven rotations at seven positions per level, this totalled 147 unique testing points.

For the first test, the tracking performance of two anatomical features of the hand were also compared to understand optimal tracking location on the hand. The anatomical features selected were the index finger's tip (distal phalanx) and knuckle (metacarpophalangeal joint). The index finger was selected as it is generally the most medial finger during interaction in VR, which provides the clearest line of sight and, therefore, the best tracking potential for the HMD cameras. The tip of the finger and knuckle were the two anatomical features on the index finger chosen within this characterization as they provide clear

⁸<https://help.vicon.com/space/Tracker39>

⁹<https://github.com/protocolbuffers/protobuf>

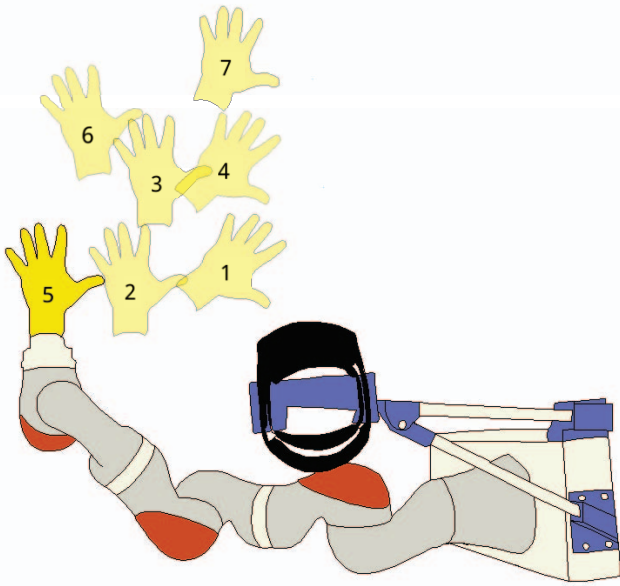


Fig. 4: Static performance testing region for each level used in Test 1.

landmarks on the hand. Features such as the palm and wrist were excluded due to the difficulty in aligning these features on the model hand to that of the virtual hand. We are unaware of any other work that compares the tracking of different anatomical features on the hand for commercial HMDs.

At each unique point, measurement was delayed for 1.5 seconds to ensure the hand-tracking latency effects were nullified. Once the delay was complete, 50 measurements were taken from both the ground truth and VR data at a sampling rate of 90 Hz (VR data transmission rate as this was the combined slowest update to ensure unique measurements at each sample). For each testing point, two results were calculated: average Euclidean error and sample-to-sample jitter.

Average Euclidean error is computed by taking the mean of the 50 measurements at each point from both the motion capture system and the markerless hand-tracking of the headset. The homogeneous transformation matrix obtained from registration is then applied to convert the positions to the same coordinate frame. The Euclidean distance between the two points can then be calculated to produce the average error. Euclidean error is an important metric for characterizing tracking performance as separation from ground truth is a true indicator of the deviation from desired performance.

Jitter represents the magnitude of frame-to-frame jumps in subsequent stationary measurements [27]. Thus, sample-to-sample jitter provides a measure of the noise experienced by the measurement system. High amounts of jitter also produce negative effects by creating less natural hand motions that can reduce social presence and make an interlocutor less comfortable in a social virtual interaction [2]. This metric is calculated by taking the root mean square (RMS) of the 50 measurements at each unique testing point. Equation 1 outlines the RMS calculation where n is the number of measurements (50 in our case) and Δm_i is the difference in samples i and $i+1$.

$$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^{n-1} \Delta m_i^2} \quad (1)$$

6.2 Test 2: Dynamic Performance

This test evaluates the hand-tracking performance under dynamic conditions to evaluate the latency of the markerless hand-tracking on each HMD. Latency refers to the time delay between ground-truth motion and when that corresponding motion is detected by the measurement

system. Since our ground truth is operating at a refresh rate of 240 Hz with a 3.6 ms camera latency, these potential delays must be accounted for in our results. The maximum latency of the ground truth (camera latency with an additional one full cycle at 240 Hz) will be 7.77 ms, which will be considered in each result in Test 2. From the user's perspective within a VR application, latency encompasses both the temporal and positional error in a dynamic interaction and, therefore, provides essential information on hand-tracking limitations. Latency is also an important factor for the embodiment of a virtual avatar and affects the ability to perform dynamic contextual interactions such as non-verbal communication [21].

12 constant-velocity trajectories of the hand are defined across three vertical levels at 0.3 m, 0.4 m, and 0.5 m below the eye level of the HMD. The testing volume was modified from Test 1 to facilitate smooth joint velocities across all trajectories. From initial testing, trajectories on vertical levels above 0.3 m below eye level would cause certain manipulator joints to have extremely rapid motion, which had greater difficulty following the desired linear trajectory. The manipulator also occluded the hand in some positions at higher vertical levels, which did not facilitate consistent tracking. Therefore, trajectories on vertical levels above 0.3 m were not included in this test. At each vertical level, four trajectories were defined: two at a distance of 0.3 m and two at a distance of 0.45 m in the anterior direction from the HMD. Out of each pair of trajectories at a unique vertical and anterior position, one moves the hand from left to right and the other from right to left. Each trajectory spanned 0.3 m and had steady-state velocities of 0.2 m/s with constant acceleration and deceleration periods lasting ten percent of the total trajectory. The acceleration and deceleration periods were established to mimic natural hand motion and to minimize the vibrations experienced by the system. During each trajectory, data was recorded from the ground-truth and VR systems at the highest update rate available (ground-truth: 240 Hz). This produced positional information from each measurement frame across the full trajectory. Data recording begins 0.25 s before the start of each trajectory and ends 1.0 s after the end, ensuring the full trajectory could be captured by both measurement systems due to the expected latency of the hand-tracking. The knuckle is selected as the tracking feature for this test (due to better tracking performance than the fingertip in Test 1). This helps minimize the effect of positional error when trying to isolate for tracking latency.

While other studies (i.e., [1, 27]) only report a single value for VR system latency, in reality, latency varies dramatically throughout different phases of motion [33]. Variable performance throughout motion can be attributed to a motion prediction step common in modern hand-tracking algorithms [11]. This variability was initially confirmed in our pilot study, as latency was highest at motion onset and became minimal at steady-state, but a more thorough analysis is required [10]. This test will, therefore, produce three metrics to provide a complete evaluation of hand-tracking performance across an entire motion: cold-start latency, motion prediction "warm-up" duration, and constant-velocity latency.

Cold-start latency refers to the time delay of the onset of motion. This measure removes the effect of motion prediction, as the hand begins the trajectory stationary, and represents the worst-case tracking condition within our tests. Since both our measurement systems experience noise, using a simple motion onset detection technique would not be viable. Warburton et al. [33] demonstrate the practicality of an outlier detection technique based on the median absolute deviation (MAD) to detect motion onset in a VR system. This technique will be used for the ground truth and VR hand-tracking data to detect the time of motion onset for each trajectory. The corresponding delay between ground truth and VR hand-tracking motion onset is our cold-start latency. The MAD detection technique is shown in Equation 2 [18].

$$\frac{x_i - M}{b \cdot M_i(|x_i - M_j(x_j)|)} > |\pm c| \quad (2)$$

This technique relies on the median function (M), constants b and c , and a window of samples (x). Leys et al. note that $c = 3$ is a conservative selection and $b = 1.4826$ is generally used as it is tied to the assumption of normality in the data [18]. These constant values will be

used in our analysis. In preliminary tests, a window size of 60 samples was shown to have strong performance. If the VR detection point was ever before the ground-truth detection (indicating a false positive as the VR hand-tracking typically succeeds ground truth motion), the window size was increased by ten until this was no longer the case. The lengthening of the window lowered the impact of high noise features within a stationary measurement to help improve detection in some cases.

As the stationary start removes the influence of motion prediction, motion prediction will begin to influence performance once the algorithm has adequate time to evaluate the motion. As noted by Warburton et al., there is a “warm-up” duration for the motion prediction to take effect and will cause the experienced latency to decrease from motion onset until a steady-state condition is reached [33]. To determine this motion prediction *warm-up duration*, latency will be determined across the entire motion first; the ground truth and VR trajectories will be aligned using the corresponding detected motion onset points from the cold-start latency analysis. From motion onset, each ground-truth measurement will be projected forward in time along the VR trajectory until a matching position is reached. The time difference between the corresponding ground-truth and VR measurements will be the latency at that specific ground-truth time. This process is repeated until 80 percent of the ground-truth trajectory is evaluated to discount the deceleration phase of the motion. With latency over most of the trajectory determined, the warm-up duration is defined as the time from motion onset for the latency to become within five percent of the final latency value within the trajectory.

Once the warm-up duration is found, the *constant-velocity latency* can then be determined by averaging latency values after the warm-up period. The period used to determine constant-velocity latency represents the best tracking condition analyzed in this test as the motion prediction is in full effect and the trajectory is outside of its acceleration phase.

6.3 Test 3: Motion-to-Photon Latency

While the first two tests provide an in-depth evaluation of hand-tracking performance across static and dynamic conditions, another key delay to consider within VR systems is the *motion-to-photon latency*. Motion-to-photon latency encompasses the delay between user motion and when that motion is displayed on the HMD to evaluate the visualization ability of the HMD. Assessing this metric is crucial, as motion-to-photon latency integrates the hardware and software performance of the HMD into a single measure. It extends the results of the second test by quantifying the tracking system’s end-to-end latency, which is the true delay experienced by the user.

To measure this latency, we have the robot travel along the same trajectories programmed for Test 2. The knuckle is again used as the anatomical feature tracked to minimize the effect of positional error when trying to isolate for tracking latency. However, we also define a “trigger point” at each trajectory’s midpoint for which both the robot’s Cartesian end-effector pose and the VR HMD’s reported hand pose are known and labelled - i.e., we have coordinates for a trigger point in both the robot and VR coordinate frames. Once the hand is seen to cross the trigger point according to the robot’s own reported pose, an LED ring on the manipulator’s wrist will be turned on. Once the VR HMD sees the hand crossing the trigger point, the entire HMD screen is coloured purple.

Similar to the procedure used by Warburton et al. in [33], a high-speed camera on a smartphone was used to record the entire setup, including the robot’s LED ring and the HMD’s screen, to determine motion-to-photon latency. In our case, an iPhone 14 Plus (iOS 17.6.1), mounted on a tripod, was used to record videos from this test at a resolution of 1080x1920 pixels and a frame rate of 240 fps. The frame rate was validated manually using a microcontroller and LED, leading us to determine that the actual frame rate was 239.69 fps; this value was used in our calculations. Following the test, videos were played back frame-by-frame from when the robot’s LED turned on to when the HMD’s screen turned purple. Counting the number of frames between these two events and dividing by the framerate yields an estimate of

motion-to-photon latency.

Note that the motion capture system was not used for this test; instead, the robot’s reporting of its kinematic state based on joint encoder feedback was used to determine the hand’s crossing of the trigger point. This method was selected because the robot’s feedback rate is virtually latency-free (within 1 ms) due to the real-time nature of the robot’s control loop. This improves our estimates of motion-to-photon latency.

7 RESULTS

7.1 Test 1: Static Performance

Data collection for Test 1 resulted in 1176 total measurements. Unfortunately, 19 of these measurements were discarded as the HMDs could not successfully track the hand. Unsuccessful tracking was deemed if the hand was no longer registered within the virtual environment or if the hand experienced abnormally high jitter while stationary.

Static positional error for both the Quest Pro and Quest 3 separated by anatomical feature conditions are displayed in Figure 5. A two-way ANOVA examined the effect of HMD and anatomical features selected on positional error. There was a statistically significant interaction between the effects of HMD and anatomical feature selection on positional error, $F(1, 1741) = 8.686$, $p = 0.003$. Simple main effect analysis showed that the Quest Pro had significantly less static positional error than the Quest 3 in their corresponding anatomical feature conditions ($p < 0.001$). Results also showed that the knuckle tracking condition on the Quest 3 experienced significantly less error than the fingertip condition ($p < 0.001$), but this was not the case between the Quest Pro tracking conditions ($p = 0.150$). Mean results for the static positional error across Test 1 conditions are displayed in Table 2.

	Static Error (cm)	Jitter (cm)
Quest Pro - Fingertip	1.30 ± 0.79	2.16 ± 0.56
Quest Pro - Knuckle	1.22 ± 0.90	1.70 ± 0.41
Quest 3 - Fingertip	2.05 ± 0.90	1.42 ± 0.72
Quest 3 - Knuckle	1.73 ± 0.72	1.11 ± 0.20

Table 2: Static positional error and jitter for each HMD and anatomical feature condition evaluated in the static test presented as an average across the testing volume.

Jitter across HMD and anatomical feature conditions are displayed in Figure 6. A two-way ANOVA examined the effect of HMD and anatomical features selected on jitter. A statistically significant interaction was not found between the effects of HMD and anatomical feature selection on jitter, $F(1, 1741) = 2.542$, $p = 0.111$. This indicates that the influence of HMD on jitter is independent of the anatomical feature tracking condition. Simple main effect analysis showed that the Quest 3 significantly outperformed the Quest Pro in corresponding anatomical feature conditions for jitter ($p < 0.001$). The knuckle was also shown to have significantly lower jitter than the fingertip condition for both HMDs ($p < 0.001$). Mean results for jitter across Test 1 conditions are displayed in Table 2.

Lastly, in Test 1, we examine the influence of hand rotation on both static positional error and jitter. A one-way ANOVA did not show statistically significant differences for jitter between the seven tested hand rotations. However, we did detect a significant difference in static error ($F(6, 434) = 2.902$, $p = 0.009$). A Tukey post hoc test for error revealed no significant difference between groups, demonstrating that the effect size may be too small to make definitive claims for this result. The average results across all positions in the testing region for each hand rotation are shown in Table 3.

7.2 Test 2: Dynamic Performance

Significant three-way interactions between headset, vertical level, and anterior proximity of linear trajectories across all results in the second test were not found. Therefore, mean values across the testing volume are reported to produce results that provide an understanding of average performance within the normal workspace of a user’s hand.

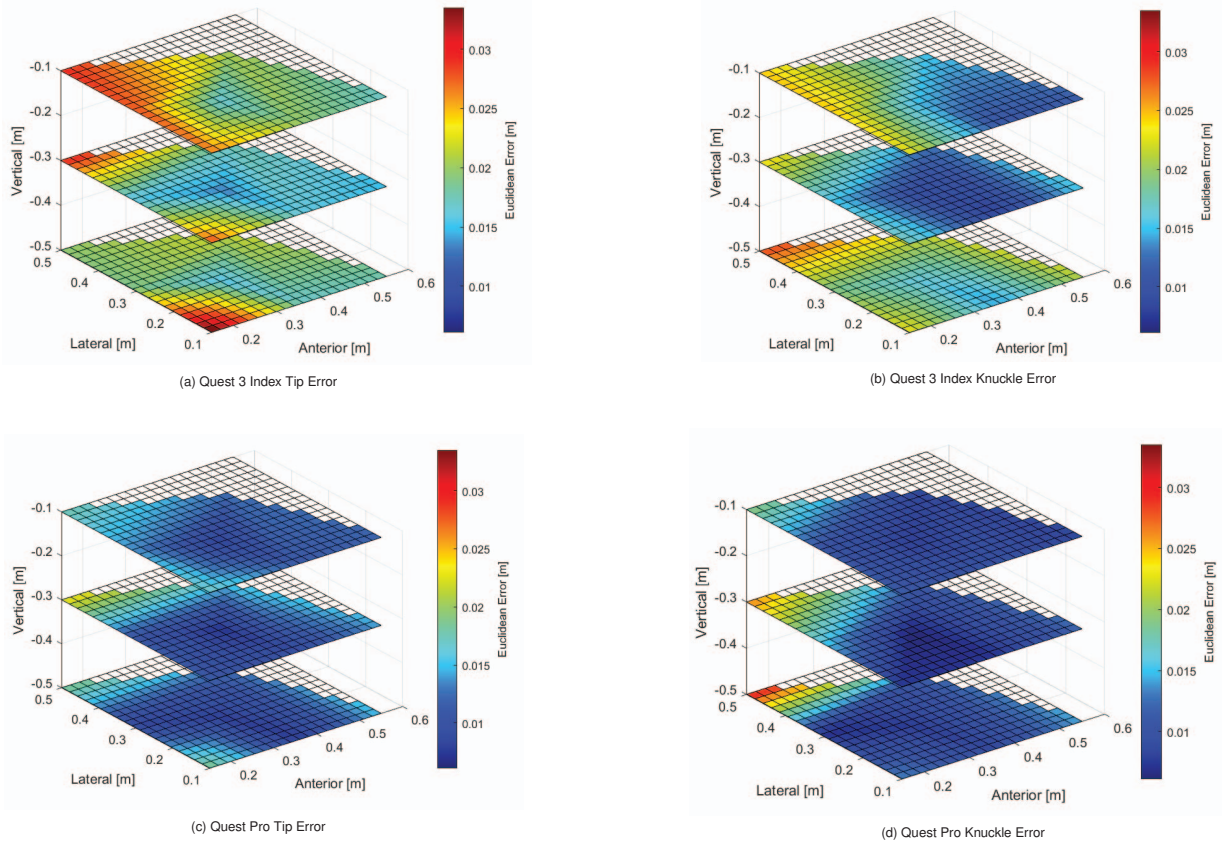


Fig. 5: Euclidean error across testing volume for each VR HMD and hand feature tracking condition from Test 1. Note that these plots linearly interpolate between positions where measurements were taken to produce the displayed spatial performance profile.

	Static Error (cm)	Jitter (cm)
Rotation 1	1.09 ± 0.98	1.66 ± 0.28
Rotation 2	1.05 ± 0.62	1.71 ± 0.32
Rotation 3	1.07 ± 0.58	1.81 ± 0.38
Rotation 4	1.18 ± 0.81	1.80 ± 0.72
Rotation 5	1.17 ± 1.00	1.77 ± 0.33
Rotation 6	1.50 ± 0.94	1.78 ± 0.31
Rotation 7	1.49 ± 1.16	1.84 ± 0.32

Table 3: Static positional error and jitter for each hand rotation evaluated in the static test presented as an average across the testing volume.

	Quest Pro	Quest 3
CS Latency (ms)	229.2 ± 146.3	220.5 ± 101.4
CV Latency (ms)	15.84 ± 9.69	14.44 ± 9.50
“Warm-Up” Duration (ms)	660.9 ± 150.5	662.0 ± 109.5
M-t-P Latency (ms)	33.20 ± 18.01	38.33 ± 31.29

Table 4: Cold-start (CS) latency, constant-velocity (CV) latency, and motion prediction warm-up duration in the dynamic test (test 2) are shown for both HMDs tested. Additionally, the motion-to-photon (M-t-P) latency from test 3 is presented. All results are presented as averages across the testing volume for the two HMDs.

Figure 7 displays an example trajectory as recorded by the Quest Pro’s markerless hand-tracking and the motion capture system (ground truth). The plots also visualize latency development across a trajectory as the motion prediction within the tracking algorithm passes its warm-up duration. A two-sided T-test failed to reveal any significant differences between average cold-start latencies, warm-up duration, and constant-velocity latencies of the Meta Quest Pro and Quest 3. These results are displayed in Table 4.

7.3 Test 3: Motion-to-Photon Latency

A two-sided t-test of motion-to-photon latencies of the Meta Quest Pro and Quest 3 did not reveal any significant differences. The motion-to-photon test results are displayed in Table 4.

8 DISCUSSION

Work by Abdalkarim et al. [1] represents, to the authors’ knowledge, the only other characterization of markerless hand-tracking on commercial HMDs. While their study provided a foundational framework for hand-

tracking characterizations, we aim to provide a more comprehensive data set. Metrics considered in [1] consider a single average value for latency. As discussed earlier and again supported by the results in this study, it appears that a single latency value is insufficient to understand how tracking performance develops over a trajectory. This motivated the inclusion of multiple evaluations for latency within our characterization, included in our second and third tests, to provide sufficient information on markerless hand-tracking latency. A benefit of the prior framework was the target-reaching task to evaluate static positional accuracy. This test evaluated the hand-tracking performance across a testing volume, which produced results that showed performance dependence on proximity to the HMD. This task provided direct motivation for developing the testing volume within this study and ensured spatial performance was properly considered within our framework.

It is clear from the results of Test 1, as shown in Figure 5 and Figure 6, that the tracking performance under all test conditions depends on the proximity to the headset. As all HMDs have cameras with fixed fields of view, identifying volumes that elicit strong tracking can help VR designers develop proper interactions in the virtual environment.

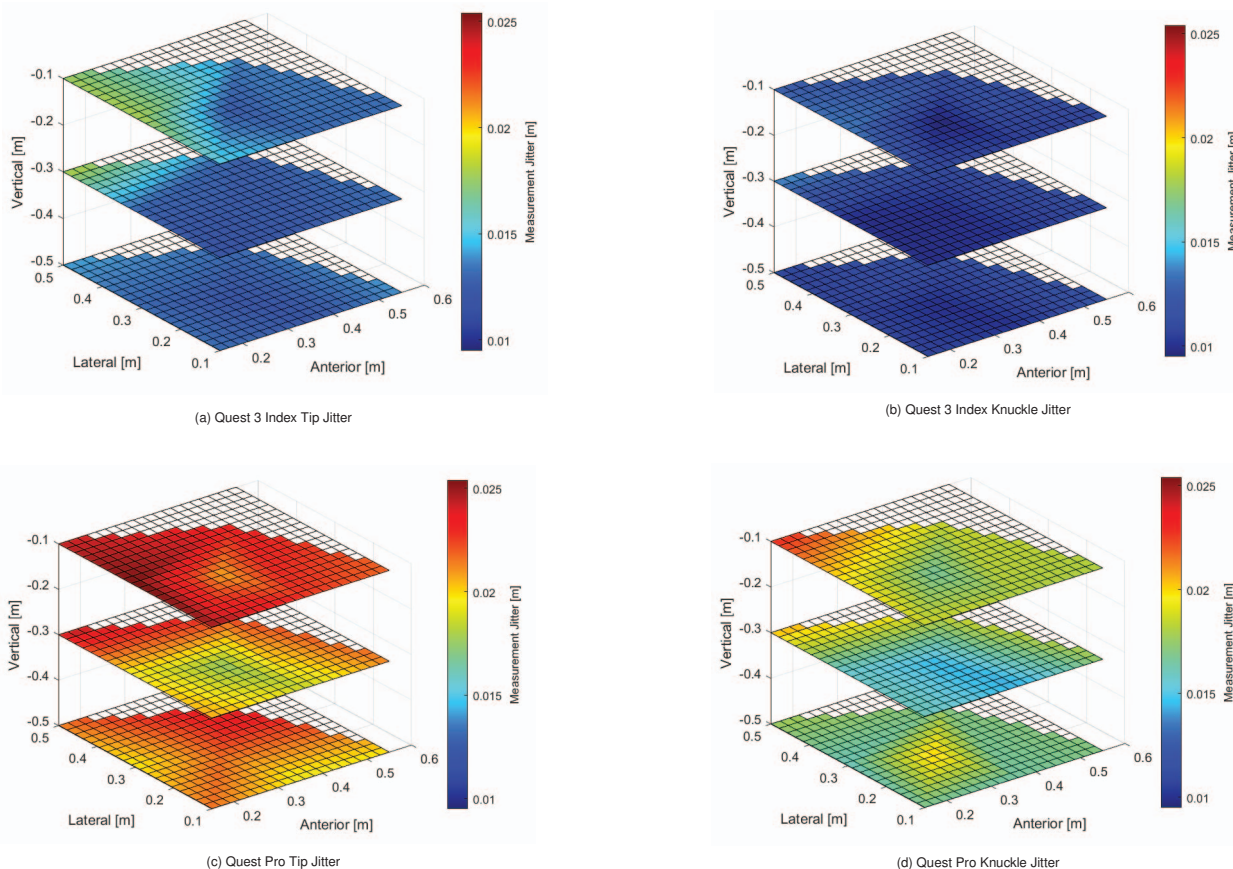


Fig. 6: Jitter across testing volume for each VR HMD and hand feature tracking condition from Test 1. Note that these plots linearly interpolate between positions where measurements were taken to produce the displayed spatial performance profile.

Position 5 (refer to Figure 4) generally delivered the worst performance as it was the furthest lateral point along the first anterior layer. This positioned the hand in an awkward position relative to the HMD and served as a blind spot that occasionally experienced drops in tracking. The best tracking was found in positions with small lateral and some vertical deviation from the headset, as in the second vertical level. This was expected due to the orientations of the cameras being positioned with a forward-facing field-of-view and the likeliness that most of the training data for the hand-tracking algorithm were captured in similar positions, as these are the positions a user's hand most often occupies. Test 1 also demonstrated the superior performance experienced in the knuckle tracking conditions compared to their corresponding fingertip conditions. As the fingertip has a higher likelihood of being victim to self-occlusions experienced by the hand, the tracking algorithm must make predictions of fingertip location with less information during these instances, which would lead to worse performance. The results align with this rationale which enforces the knuckle as the superior tracking location out of the two anatomical features tested. While accurate fingertip tracking is crucial for tasks like fingertip-based user interactions in VR [6], this finding underscores the knuckle's greater reliability as a tracking feature, making it preferable for general purposes such as hand pose identification.

Combining the result of this study with the work of Abdulkarim et al. in [1], the hand-tracking performance of the past three iterations of Meta Quest HMDs can be compared. The average static fingertip positional error of the Meta Quest 2 was shown to be 1.1cm [1]. The average results for our fingertip conditions were 2.05cm and 1.30cm for the Meta Quest 3 and the Meta Quest Pro, respectively. While our result shows a greater average at face value, Test 1 records measurements while the hand is at 7 different rotations for each position within the testing volume. By rotating the hand out of the standard palm-down

configuration, which was standard for the Meta Quest 2 evaluation, the tracking difficulty increases due to the potential for self-occlusions and/or reduced surface area from the HMD cameras' perspective. The results within this study present a better evaluation of general tracking performance as both spatial and rotational conditions were considered. With respect to the rotational conditions, performance was best when the hand was in a palm-down configuration (Rotations 1-3). The pronation and supination conditions of Rotation 6 and Rotation 7 produced the worst performance. This was likely due to the hand's surface area drastically decreasing from the perspective of the HMD as the hand rotated along its anterior axis. As natural hand interactions rarely occur in a palm-down configuration, decreased tracking performance should be accounted for when designing interactions that differ from this ideal state.

Initially proposed by Warburton et al. in [33], our results confirm that multiple latency results are required to fully evaluate tracking latency. This is evident by the dramatic difference between the cold-start and constant-velocity latency values. The cold-start latency results evaluate tracking performance under worst-case conditions as the motion prediction effects of the tracking algorithm are mitigated. The average cold-start latency results of 229.2ms and 220.5ms for the Meta Quest Pro and Quest 3, respectively, are above the previously mentioned discernible thresholds for visual latency as measured in [15, 17]. This demonstrates that the starting portion of trajectories, or hand motions with abrupt accelerations, can experience latencies that the user may detect. This could affect user experience, and hand motions with low accelerations should be desired when designing hand interactions within a virtual experience. The constant-velocity results drop below the mentioned discernible visual thresholds for latency which demonstrates the importance of motion prediction within the hand-tracking algorithm on Meta Quest HMDs. The warm-up duration for motion prediction

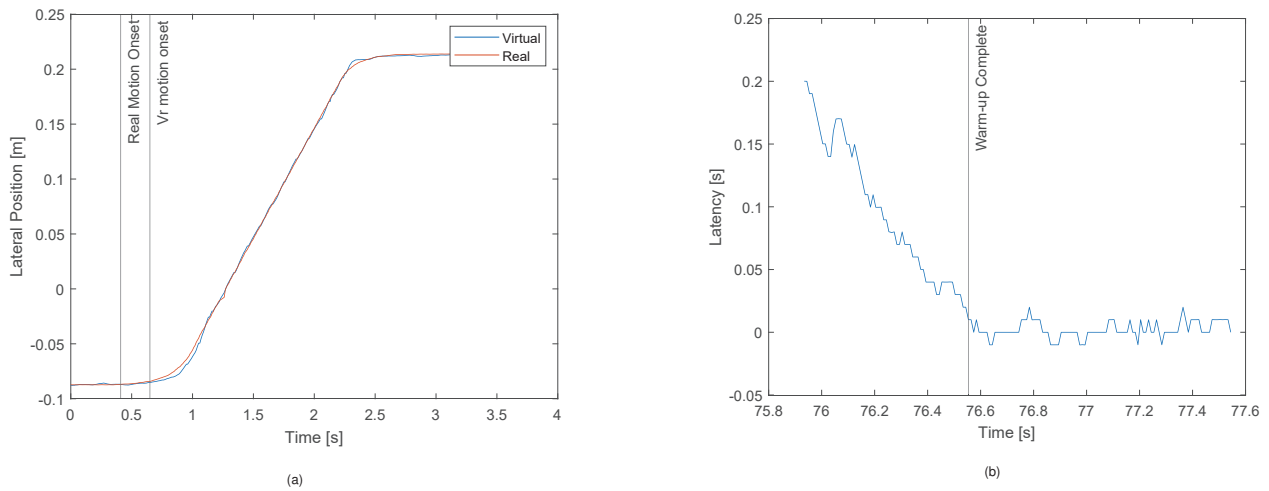


Fig. 7: Example output plots for Test 2 - Dynamic Testing showing: (a) VR hand-tracking vs. ground-truth across linear trajectory with marked motion onset points used for cold-start latency analysis; and (b) development of tracking performance across linear trajectory.

is a key result to understand how quickly peak tracking performance can be reached, as during constant velocity motion. This warm-up period represents the transient state of tracking performance and, again, must be considered when relying on markerless hand-tracking data for applications requiring precise tracking. Since the warm-up duration results for both HMDs did not show significantly different results, it is likely Meta is using the same predictive algorithm on each HMD.

The motion-to-photon latency results of the final test represent the best indicator of complete system performance within this study as it encompasses both the hand-tracking and visualization performance on the HMDs. The results of 33.20ms and 38.33ms for the Meta Quest Pro and the Meta Quest 3, respectively, are below the discernible thresholds of 50ms [17] and 90ms [15] discussed before. It must be noted that the 50ms threshold for visual perception is breached for the results of both headsets when one standard deviation is considered. This demonstrates that although the average performance of markerless hand-tracking on Meta HMDs may be sufficient to enable a positive user experience, discernible effects produced by system latency will be common, which may have negative implications.

While this work aims to provide an extensive framework for markerless hand-tracking characterization, limitations present opportunities for future work. Across the three tests, the testing volume only considered positions and trajectories located below the HMD. As most hand interactions occur below the eyeline, the testing volume was defined to best match those locations. As virtual experiences become more complex and potentially demand more from markerless hand-tracking on VR systems, characterization across a larger volume may be required to facilitate new interactions. Also, the test conditions within this work only involved single joint tracking and did not evaluate the tracking of more macro features, such as finger angles or handbreadth. This evaluation is also limited to a single static hand model. This limitation may affect the generalizability of the results, and future studies could explore variations in hand shapes to address this issue. Many additional tracking conditions exist that would increase the depth of analysis provided by the results of this study.

9 CONCLUSION

Markerless hand-tracking represents a key development pathway for improving future virtual experiences. This technology, currently being standard on many commercial HMDs, demonstrates a push to facilitate virtual interaction with natural hand manipulations. While implementing hand-tracking in virtual experiences is an important step forward, consistent and ongoing evaluations are necessary to properly characterize and document the progress of this technology for applications

requiring accurate and precise hand pose data. This has motivated us to develop a repeatable and extensible process of markerless hand-tracking characterization within this work to not only facilitate the evaluation of current commercial HMDs but also to lay the foundation for future testing and comparison.

Our work represents, to our knowledge, the first use of robotics to characterize markerless hand-tracking performance. Doing so allows for systematic, precise, and comparable characterization of markerless hand-tracking on VR systems. The methodology outlined in this work produces results of positional accuracy, jitter, and latency. This methodology was applied to two commercially available VR HMDs: the Meta Quest Pro and Quest 3 HMDs. In best-performing conditions (tracking knuckle), the Quest Pro outperformed the Quest 3 with respect to positional error (1.22 cm and 1.73 cm, respectively). The opposite was true for jitter (1.77 cm and 1.11 cm, respectively). The effect of hand proximity to HMD on positional accuracy and jitter was evaluated, and weaker performance was observed as the hand moved to the periphery for both metrics. Latency results were also highly variable for the Quest Pro (15.8 - 229.2 ms) and Quest 3 (14.4 - 220.5 ms).

While the characterization framework developed in this work is comprehensive, the analysis was completed across a limited volume. Only points below the eye-line of the HMD were tested, and constraints imposed by the manipulator limited the overall testing workspace. Future work can grow the knowledge developed within this work by expanding the testing volume of future characterizations and evaluating the markerless hand-tracking performance of other commercial VR HMDs.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada and the Connected Minds CFREF program (<https://www.yorku.ca/research/connected-minds/>).

REFERENCES

- [1] D. Abdikarim, M. Di Luca, P. Aves, M. Maaroufi, S.-H. Yeo, R. C. Miall, P. Holland, and J. M. Galea. A methodological framework to assess the accuracy of virtual reality hand-tracking systems: A case study with the meta quest 2. *Behavior research methods*, 56(2):1052–1063, 2024. 2, 4, 5, 7, 8
- [2] A. Adkins, A. Normoyle, L. Lin, Y. Sun, Y. Ye, M. Di Luca, and S. Joerg. How important are detailed hand motions for communication for a virtual character through the lens of charades? *ACM Transactions on Graphics*, 42(3):1–16, 2023. 5
- [3] T. Ameler, K. Blohme, L. Brandt, R. Brüngel, A. Hensel, L. Huber, F. Kuiper, J. Swoboda, M. Warnecke, M. Warzecha, et al. A comparative

- evaluation of steamvr tracking and the optitrack system for medical device tracking. In *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pp. 1465–1470. IEEE, 2019. 2
- [4] G. Buckingham. Hand tracking for immersive virtual reality: opportunities and challenges. *Frontiers in Virtual Reality*, 2, 2021. 1
- [5] P. Caserman, M. Martinussen, and S. Göbel. Effects of end-to-end latency on user experience and performance in immersive virtual reality applications. In *Entertainment Computing and Serious Games: First IFIP TC 14 Joint International Conference, ICEC-JCSG 2019, Arequipa, Peru, November 11–15, 2019, Proceedings 1*, pp. 57–69. Springer, 2019. 2
- [6] M. Dani, G. Garg, R. Perla, and R. Hebbalaguppe. Mid-air fingertip-based user interaction in mixed reality. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 174–178, 2018. doi: 10.1109/ISMAR-Adjunct.2018.00061 8
- [7] D. Eger Passos and B. Jung. Measuring the accuracy of inside-out tracking in xr devices using a high-precision robotic arm. In *International Conference on Human-Computer Interaction*, pp. 19–26. Springer, 2020. 2
- [8] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly. Vision-based hand pose estimation: A review. *Computer Vision and Image Understanding*, 108(1-2):52–73, 2007. 2
- [9] Y. Ferstl, R. McDonnell, and M. Neff. Evaluating study design and strategies for mitigating the impact of hand tracking loss. In *ACM Symposium on Applied Perception 2021*, pp. 1–12, 2021. 2
- [10] E. Godden and M. Pan. Towards robotic haptic proxies in virtual reality. In *2024 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp. 178–179, 2024. doi: 10.1109/CCECE59415.2024.10667136 5
- [11] S. Han, B. Liu, R. Cabezas, C. D. Twigg, P. Zhang, J. Petkau, T.-H. Yu, C.-J. Tai, M. Akbay, Z. Wang, et al. Megatrack: monochrome egocentric articulated hand-tracking for virtual reality. *ACM Transactions on Graphics (ToG)*, 39(4):87–1, 2020. 2, 5
- [12] J. Hesch, A. Kozminski, and O. Linde. Powered by ai: Oculus insight, Aug 2019. 2
- [13] V. Holzwarth, J. Gisler, C. Hirt, and A. Kunz. Comparing the accuracy and precision of steamvr tracking 2.0 and oculus quest 2 in a room scale setup. In *Proceedings of the 2021 5th International Conference on Virtual and Augmented Reality Simulations*, pp. 42–46, 2021. 2
- [14] M. Huber, C. E. Mower, S. Ourselin, T. Vercauteren, and C. Bergeles. Lbr-stack: Ros 2 and python integration of kuka fri for med and iiwa robots, 2023. 3
- [15] C. Jay and R. Hubbard. Delayed visual and haptic feedback in a reciprocal tapping task. In *First joint eurohaptics conference and symposium on haptic interfaces for virtual environment and teleoperator systems. World Haptics Conference*, pp. 655–656. IEEE, 2005. 2, 8, 9
- [16] T. A. Jost, G. Drewelow, S. Koziol, and J. Rylander. A quantitative method for evaluation of 6 degree of freedom virtual reality systems. *Journal of biomechanics*, 97:109379, 2019. 2
- [17] I. Lee and S. Choi. Discrimination of visual and haptic rendering delays in networked environments. *International Journal of Control, Automation and Systems*, 7(1):25–31, 2009. 2, 8, 9
- [18] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of experimental social psychology*, 49(4):764–766, 2013. 5
- [19] G. Lugo, M. Ibarra-Manzano, F. Ba, and I. Cheng. Virtual reality and hand tracking system as a medical tool to evaluate patients with parkinson's. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pp. 405–408, 2017. 1
- [20] I. S. MacKenzie and C. Ware. Lag as a determinant of human performance in interactive systems. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pp. 488–493, 1993. 2
- [21] D. Maloney, G. Freeman, and D. Y. Wohn. "talking without a voice" understanding non-verbal communication in social virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–25, 2020. 5
- [22] M. Mangalam, S. Oruganti, G. Buckingham, and C. W. Borst. Enhancing hand-object interactions in virtual reality for precision manual tasks. *Virtual Reality*, 28(4):166, 2024. 4
- [23] A. Mc Kittrick, M. R. Desselle, A. Padilha Lanari Bo, B. Zhang, S. Laracy, and G. Tornatore. Patient Experience in Adjunct Controller-Free Hand Tracking Virtual Reality Tasks for Upper-Limb Occupational Therapy Rehabilitation. *J Patient Exp*, 10:23743735231211983, 2023. doi: 10.1177/23743735231211983 1
- [24] C. Mizera, T. Delrieu, V. Weistroffer, C. Andriot, A. Decatoire, and J.-P. Gazeau. Evaluation of hand-tracking systems in teleoperation and virtual dexterous manipulation. *IEEE Sensors Journal*, 20(3):1642–1655, 2019. 2
- [25] C. Mizera, T. Delrieu, V. Weistroffer, C. Andriot, A. Decatoire, and J.-P. Gazeau. Evaluation of Hand-Tracking Systems in Teleoperation and Virtual Dexterous Manipulation. *IEEE Sensors Journal*, 20(3):1642–1655, Feb. 2020. Conference Name: IEEE Sensors Journal. doi: 10.1109/JSEN.2019.2947612 1
- [26] F. Mueller, D. Mehta, O. Sotnychenko, S. Sridhar, D. Casas, and C. Theobalt. Real-time hand tracking under occlusion from an egocentric rgb-d sensor. In *Proceedings of the IEEE international conference on computer vision*, pp. 1154–1163, 2017. 2
- [27] D. C. Niehorster, L. Li, and M. Lappe. The accuracy and precision of position and orientation tracking in the htc vive virtual reality system for scientific research. *i-Perception*, 8(3):2041669517708205, 2017. 2, 5
- [28] M. K. Pan and G. Niemeyer. Catching a real ball in virtual reality. In *Proceedings - IEEE Virtual Reality*, pp. 269–270. IEEE Computer Society, Apr. 2017. doi: 10.1109/VR.2017.7892280 1
- [29] Y. Sato, Y. Kobayashi, and H. Koike. Fast tracking of hands and fingertips in infrared images for augmented desk interface. In *Proceedings fourth ieee international conference on automatic face and gesture recognition (cat. no. pr00580)*, pp. 462–467. IEEE, 2000. 2
- [30] G. Schreiber, A. Stemmer, and R. Bischoff. The fast research interface for the kuka lightweight robot. In *Proc. IEEE Int. Conf. Robotics and Auto.*, pp. 15–21, 2010. 3
- [31] E. Theodoridou, L. Cinque, F. Mignosi, G. Placidi, M. Polsinelli, J. M. R. Tavares, and M. Spezialetti. Hand tracking and gesture recognition by multiple contactless sensors: A survey. *IEEE Transactions on Human-Machine Systems*, 53(1):35–43, 2022. 2
- [32] T. Waltemate, I. Senna, F. Hülsmann, M. Rohde, S. Kopp, M. Ernst, and M. Botsch. The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality. In *Proceedings of the 22nd ACM conference on virtual reality software and technology*, pp. 27–35, 2016. 1, 2
- [33] M. Warburton, M. Mon-Williams, F. Mushtaq, and J. R. Morehead. Measuring motion-to-photon latency for sensorimotor experiments with virtual reality systems. *Behavior research methods*, 55(7):3658–3678, 2023. 5, 6, 8