TECHNICAL EVALUATION OF THE FIDELITY OF THE HTC VIVE FOR UPPER LIMB TRACKING

Matteo Mancuso[1*], Caecilia Charbonnier[1, 2]

[1]Artanim Foundation,
[2]Faculty of Medicine, University of Geneva (UNIGE)
*Correspondence: matteo.mancuso@artanim.ch

Novel tracking technologies targeting non-technical users enabled the creation of accessible motion tracking for medical functional evaluation and assisted rehabilitation. Here we describe the adaptation of the HTC Vive tracking technology toward this goal. The resulting system was then tested comparing three different anatomical calibration methods, accounting for inter-subject (1 female, 2 male) and inter-operator (2 experienced, 3 inexperienced) variability, as well as six different methods for spatial alignment with the reference measurements system. Accuracy and precision of our methods were quantified by comparison with Vicon optical tracking system. The accuracy of the anatomical landmark tracking was 2.3 ± 0.5 cm with the cleanest protocol tested. Inter-operator variability caused significant changes in measurement errors, while inter-subject variability did not.

KEYWORDS: motion capture, virtual reality, anatomical calibration, functional assessment, rehabilitation, accuracy.

INTRODUCTION: The advent of virtual reality (VR) technologies has caused a surge of interest in simple motion tracking technologies for home usage. This led to the adaptation of complex and expensive research-grade motion tracking technologies into simpler and cheaper solutions, which are either camera-based, where a few cameras are positioned around the subject (i.e., outside-in tracking) or embedded in the VR headset (i.e., inside-out tracking). This study focused on the second generation of the Lighthouse tracking technology (LTT), provided with the HTC Vive and its *tracked objects* (hand controllers and Vive trackers). LTT has been commercially available since 2016 (1st generation) and is used in some large communities such as *VRChat* (VRChat Inc., n.d.) as the best way to achieve full body tracking at home. Its positional accuracy was evaluated to be around 2 mm (Kreylos, 2016), making it the best VR tracking solution available for home usage at the time. Additionally, the large space tracked by the sensors (5 m range from each base-station (HTC, n.d.)) made the system interesting for various use-cases including rehabilitation (Chen et al., 2018), diagnostics (Mancuso, 2020), sport analysis (Merker et al., 2023), psychological simulations (Niehorster et al., 2017), etc.
(Niehorster et al., 2017) observed two major flaws in the LTT tracking. The ground plane estimated by LTT generally presented an angular offset of a few degrees with respect to the real floor of the room, and this offset would change every time the headset would lose and recover tracking, making a calibration of the offset unpractical. The 2nd generation of LLT (2018), presented an increase in (i) tracking range of each base-station, (ii) allowed the use of an unlimited number of base-stations, and (iii) increased product's longevity. HTC did not provide any accuracy or precision specifications for this 2nd generation, and to the best of our knowledge only Bauer et al. (Bauer et al., 2021) conducted a thorough investigation on the topic, investigating tracking performance mostly in static lab's conditions. They reported an accuracy in the millimetre range, but that the system seemed to mostly rely on the location provided by one of the base-stations termed "master base-station" and that the identity of the "master base-station" as well as the ground-plane orientation could change during the experiments and cause centimetre-level errors. Merker et al. (Merker et al., 2023) used the HTC Vive trackers 3.0 to track the lower body of a subject on a bicycle ergometer and recorded centimetre level errors.
No study so far evaluated the ability of LTT to track the motions of users' upper bodies moving in space. For medical and sportive range of motion analysis, it is important to evaluate the positions of specific standardized locations on the body (Wu et al., 2005), usually defined as

palpable bony landmarks. Those landmarks are then used to build the reference frame of the different segments (arm, forearm, etc.) and to compute the joint ranges of motion (rotations and translations) in an intraparticipant and interparticipant repeatable way, making results clinically relevant and comparable with previous research.

Our goal was to evaluate this procedure by tagging (using the hand controller as a manual pointer to indicate the precise 3D position of the given landmark, to then track its relative position with respect to a tracked object placed on the same anatomical segment) a set of palpable surface markers, very close to the bone, and as reproducible as possible. In these circumstances, a palpation error around 15 mm is typically expected (Rabuffetti et al., 2002).

METHODS: Our setup was composed of 4 Lighthouse 2.0 base-stations, 1 HTC Vive Pro as a head-mounted display (HMD), 2 hand controllers 2.0, 5 Vive trackers 2.0 (HTC Corporation, Xindian, Taiwan) and a subject-specific rigid body model of the subject's joints, created through a minute-long anatomical calibration process. This process allowed us to model the relative position of the main standardized anatomical landmarks (Figure 1) with respect to the poses of the LTT trackers. Next, the LTT measurements were compared against a state-of-the-art Vicon optical tracking system (Vicon, Oxford Metrics, UK) used as a ground-truth. It is composed of 24 Vicon MXT40S tracking a capture space of 5 m wide and 10 m long. Besides assessing the accuracy and precision of this tracking solution for the anatomical landmarks of the upper limb (UL), this investigation also evaluated different sources of errors and experimentally compared the impact of the following confounding factors:

1. An anatomical calibration performed by the user himself/herself against a calibration performed by an assisting operator (36 measurements per condition).
2. A "careless" operator against a "careful" operator trying to the best of his abilities to avoid causing occlusions on the tracked objects during anatomical calibration (120 measurements per condition).
3. Inter-subject variations (2 males, 1 female) in accuracy and precision during anatomical calibration (60 measurements per subject). The subjects were selected to have as diverse body proportions as possible (height: 1.6-2.0 m; weight: 67-130 kg).
4. Inter-operator variations (2 with years of experience, 3 inexperienced) in accuracy and precision during anatomical calibration (60 measurements per operator).
5. Six protocols to spatially align the LTT reference frame with Vicon's reference frame while minimizing the creation of tracking artifact (360 measurements in total). They were divided into single frame recordings or multi-frames (10 seconds) recordings, before, after or synchronous with the test, directly using the sensors worn by the subject for the duration of the test.

Figure 1: Fully equipped subject with the HMD, a controller in each hand and the trackers (orange) strapped to pelvis, upper arms, and forearms. The image also shows the Vicon's optical markers placed on anatomical landmarks (blue).

Accuracy was computed as the difference between the position measured through LTT and the position recorded by Vicon. Precision was described by the standard deviation (SD) of LTT measurements. Descriptive statistics were presented in terms of maximum, minimum, mean and SD. The Shapiro–Wilk test was used to screen the outcomes for normal distribution. Paired samples Student T/Wilcoxon tests were applied to check for significant differences in operator's anatomical calibration techniques and ground calibration methods. Friedman's Anova's tests were used to check the inter-subject and inter-operator variations. The level of significance was set at $p < 0.05$.

RESULTS: Different operator types ($\mu_{self\text{-}calibration} = 1.66 \pm 0.85$ cm, $\mu_{careless} = 0.86 \pm 0.58$ cm, $\mu_{careful} = 1.13 \pm 0.44$ cm) indicated a significantly poorer quality of calibration for the self-calibration method compared to an assisted calibration (Wilcoxon test: $p << 0.05$), leading us to exclude this method from further tests. Careful operator displayed larger average errors (1.13 cm) than the careless one (0.86 cm). The sample-size was small and the difference non-significant, therefore a second recording session was added for further investigation (360 new measurements per condition). Combining the measurements from both sessions ($\mu_{careless}$ = 0.54 ± 0.48 cm, $\mu_{careful} = 0.59 \pm 0.55$ cm) indicated a non-significant ($p = 0.23$) difference between conditions, suggesting that the LTT tracking system seems robust to occlusions and that the operator does not need to pay particular attention to avoid optical occlusions.
Varying subject morphologies requested minor adaptation on marker's placement, but overall, no major difference in tracking quality ($\mu_{subject1} = 1.9 \pm 0.53$ cm, $\mu_{subject2} = 1.96 \pm 0.51$ cm, $\mu_{subject3}$ = 1.86 ± 0.54 cm) was observed between the different subjects (one-way Anova: $p > 0.55$), indicating a good adaptability of the system to different morphologies.
Five operators measuring in turn the same subject five times lead to different results. *Operator 3* failed to understand and perform the measurements correctly. The remaining four operators presented statistically different results, indicating the importance of the operator on the results quality (Friedman's Anova: $p = 0.046$), but comparing the experienced (*Operators 4 and 5*) to the unexperienced (*Operators 1 and 2*) presented no statistical difference ($p = 0.74$).
Spatial alignment: Since LTT has been reported to occasionally change the orientation of its ground plane, the recording of the alignment data before or after the experiment could be a cause of error. For single-frame alignment methods, recording synchronously with the anatomical calibration seemed to improve the measurements ($\mu_{synchronous} = 2.5 \pm 1.2$ cm, $\mu_{pre\text{-}recorded} = 3.5 \pm 4.1$ cm, $\mu_{post\text{-}recorded} = 2.6 \pm 1.3$ cm), but no statistical difference was found (Wilcoxon test: $p > 0.1$). The multi-frame synchronous approach indicated both improved mean (2.3 cm) and SD (0.5 cm), with respect to any of the previous methods.

DISCUSSION: Testing for operator's type allowed to clearly identify that self-calibration should be avoided but that the system is robust to occlusions and that the operator does not need to be mindful about avoiding them. However, comparing different operators led to significantly different results, which were not related to their experience tagging anatomical landmarks. They performed 5 times 12 measurements, and no learning effect was observed during repetitions. In contrast, fixing the operator and varying subject did not impact measurement quality, even while varying subject's height (1.6-2.0 m), mass (67-130 kg) or sex ($p > 0.55$).
The spatial alignment of the LTT's reference frame with Vicon's reference frame was an important factor to evaluate, as it could introduce additional bias cumulative to the anatomical calibration errors. Several options were tested, and it appeared necessary to perform the alignment synchronously with the recording in order to avoid as much as possible the errors induced by a tilt in the ground-plane. Recording on multiple frames was also reported as beneficial, as occasional sensor drifts can cause large deviations in the spatial alignment ruining the subsequent dataset. The presence of the HMD being always close to the tracked points was observed to improve the tracking quality of any tracked object in LTT, probably due to an optimisation of the tracked object active only while it is in the field of view of the HMD, or the use of the HMD as additional information to evaluate its location.
In future works, it would be interesting to study solutions to prevent or compensate the changes in the definition of the ground plane as automatically performed by the LTT system. Maintaining the subject always in front of the same base station within a relatively short range was not studied here but might also prevent the redefinition of the master base-station and the associated tilting of the floor plane.

The evaluation of the LTT system for classic kinematic analysis should be studied by evaluating the resulting joint's kinematics from the anatomical landmarks and how reliable this technology could be for the sportive and medical evaluation of musculoskeletal disorders.

CONCLUSION: This study evaluated the ability of LTT to be used for the accurate tracking of human anatomical landmarks. It tested different set-up strategies and verified that a calibration based on self-tagging was significantly worse than a calibration performed in a subject-operator configuration. Surprisingly the inter-operator variability was high, independently from the level of experience of the operators. The system was shown robust to occlusions as operators paying attention to avoid them recorded the same accuracy as those that did not. Inter-subject tracking quality was tested on 3 very different subjects but no significant effect of height, body mass or gender were observed. Due to the complexity of the tracking strategy and the lack of control on the latter, this study was able with its best setup to track anatomical landmarks with an accuracy of 2.3 ± 0.5 cm, which is significantly worse than the few millimetres observed in lab conditions (Bauer et al., 2021; Kreylos, 2016).
Overall LTT is capable technology, offering an affordable and integrated solution for complex tracking applications in VR. Its precision and accuracy in ideal conditions can go down to a few millimetres and mounting virtualized rigid body assemblies on top of the trackers can greatly extend the range of applications of the device. Unfortunately, the use of complex and undisclosed hypothesis-based tracking optimisations, make its use difficult in conditions that require consistent and reliable measurements.

REFERENCES

Bauer, P., Lienhart, W., & Jost, S. (2021). Accuracy investigation of the Pose determination of the HTC Vive. *Sensors*, *21*. https://doi.org/https://doi.org/10.3390/s21051622

Chen, D., Liu, H., & Ren, Z. (2018). Application of Wearable Device HTC VIVE in Upper Limb Rehabilitation Training. *Proceedings of 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, IMCEC 2018*, *Imcec*, 1460–1464. https://doi.org/10.1109/IMCEC.2018.8469540

HTC. (n.d.). *Vive User Guide*. 48. https://www.manualslib.com/manual/1543807/Htc-Vive.html?page=2#manual

Kreylos, O. (2016). *Lighthouse tracking examined*. Doc-Ok.Org. http://doc-ok.org/?p=1478

Mancuso, M. (2020). *Evaluation and robotic simulation of the glenohumeral joint*. EPFL.

Merker, S., Pastel, S., Bürger, D., Schwadtke, A., & Witte, K. (2023). Measurement Accuracy of the HTC VIVE Tracker 3.0 Compared to Vicon System for Generating Valid Positional Feedback in Virtual Reality. *Sensors*, *23*(17). https://doi.org/10.3390/s23177371

Niehorster, D. C., Li, L., & Lappe, M. (2017). The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research. *I-Perception*, *8*(3), 1–23. https://doi.org/10.1177/2041669517708205

Rabuffetti, M., Baroni, G., Ferrarin, M., Ferrigno, G., & Pedotti, A. (2002). Self-marking of anatomical landmarks for on-orbit experimental motion analysis compared to expert direct marking. *Human Mov Sci*, *21*, 439–455.

VRChat Inc. (n.d.). *VRChat*. https://hello.vrchat.com/

Wu, G., Van Der Helm, F. C. T., Veeger, H. E. J., Makhsous, M., Van Roy, P., Anglin, C., Nagels, J., Karduna, A. R., McQuade, K., Wang, X., Werner, F. W., & Buchholz, B. (2005). ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion - Part II: Shoulder, elbow, wrist and hand. *Journal of Biomechanics*, *38*(5), 981–992. https://doi.org/10.1016/j.jbiomech.2004.05.042

ACKNOWLEDGMENTS AND FUNDING: The authors would like to thank Bart Kevelham for his valuable help to implement, structure and debug the software used for this experiment, as well as for his help during the experimental phase. Additionally, the authors would like to thank Pierre Nagorny for his precious help while running the experiments. This work was entirely funded by Artanim Foundation's internal funds.