Tracking a Consumer HMD with a Third Party Motion Capture System

Henrique Galvan Debarba*[1] Marcelo Elias de Oliveira [1] Alexandre Lädermann [2,3,4]
Sylvain Chagué [1] Caecilia Charbonnier [1]

[1] Artanim Foundation, Meyrin, Switzerland
[2] Faculty of Medicine, University of Geneva, Geneva, Switzerland
[3] Division of Orthopedics and Trauma Surgery, La Tour Hospital, Meyrin, Switzerland
[4] Division of Orthopaedics and Trauma Surgery, Clinique La Colline, Geneva, Switzerland

Abstract

We describe a calibration procedure to track consumer Head Mounted Displays (HMD) using a 3rd party tracking solution. The calibration consists of registering the center of projection of the rendering hardware to a 3rd party tracked object attached to it, and is performed by matching motion datasets from the HMD built-in and 3rd party tracking solutions. We demonstrate this calibration with an augmented reality optical see-through HMD, where the correctness of the alignment is critical to the visual match of a real object by a virtual overlay. We assessed a mean error of $3mm$ ($SD = 1mm$) for objects at a distance of $70cm$ in the projected overlay image.

Index Terms: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality

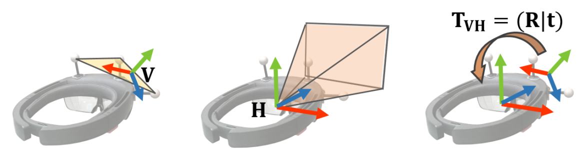
1 Introduction

With the introduction of consumer Head Mounted Displays (HMD) for virtual and augmented reality (VR/AR), low cost technologies for absolute pose tracking of these devices have also entered the market. In general, these tracking methods rely on input that is less accurate (e.g. single camera and the structural analysis of the environment) than high-end optical tracking solutions such as VICON[1] or PhaseSpace[2] systems to deduce the headset pose. Therefore, in many instances it is desirable to replace the built-in tracking with a 3rd party tracking solution. However, to do so we are required to accurately calibrate the arbitrary rigid body defined by the 3rd party tracking system to the built-in tracking of the HMD. We describe a procedure to perform this calibration by aligning datasets from the built-in and the 3rd party tracking solutions. Our calibration is performed in two steps: first, we find the rotation that aligns the axis of movement of the HMD in both tracking coordinate systems; second, we find a point common to both coordinate systems, allowing to compute the rigid transformation from the 3rd party tracking to the built-in tracking in the coordinate system of the former.

We demonstrate this calibration with a Microsoft HoloLens[3] Optical See-through (OST) HMD. The HoloLens includes optical and inertial sensors for position and orientation tracking. Although the algorithm combining the sensors information in the HoloLens can yield estimations of the headset's pose in an absolute frame of reference, it may present several centimeters of error, it cannot track additional objects, and it is not appropriate for non-static backgrounds. The AR headset tracking is combined in this paper with a VICON MXT40S motion capture system. The VICON tracking system consists of 24 infrared cameras sampling at 240 Hz and is used to track 5 retro-reflective markers (⊘ 19 mm) attached to the HoloLens.

*e-mail: henrique.debarba@artanim.ch

[1] www.vicon.com
[2] www.phasespace.com
[3] www.microsoft.com/en-us/HoloLens



(a) (b) (c)

Figure 1: (a) Coordinate system arbitrarily defined by optical markers rigidly attached to the headset. (b) Coordinate system of the headset aligned with camera projection. (c) The calibration determines the rigid transformation – rotation **R** and translation **t** – that aligns the tracked coordinate system to the headset's coordinate system.

With a set of markers placed in the HoloLens headset the VICON tracking can define an arbitrary coordinate system (Fig. 1a) with a position and orientation that do not match that of the AR headset image projection center (Fig. 1b). In order to correctly display the AR overlay in the headset, the rigid transformation that aligns the markers-based arbitrary coordinate system to the physical center of projection of the AR headset is required (Fig. 1c). High precision and accuracy are necessary, as small errors will affect the quality of the overlap of holographic and real objects. Note that the rendering with an AR or VR headset requires a pair of cameras (stereo rendering pair), this is not relevant for the calibration as the pose of both cameras can be defined relative to the coordinate system shown in Fig. 1b. For instance, by laterally translating the cameras by half of the interpupillary distance of the user (HoloLens provides a tool to estimate this distance[4]).

Different calibration methods exist for OST HMD calibration, normally relying on visual alignment of objects of 3D to 2D (e.g. Single-Point Active Alignment Method (SPAAM) [5]) or 3D to 3D correspondences [4]. Unlike related AR calibration methods [4,5], our procedure aligns the HoloLens built-in tracking and the 3rd party tracking based on classical methods [2] and it can also be used for consumer VR HMD devices.

2 Calibration Procedure

The mapping that express the center of projection **H** relative to the coordinate system **V** illustrated in Fig. 1 requires a rigid transformation with rotation **R** and translation **t**:

$$\mathbf{T_{VH}} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0^T & 1 \end{bmatrix} \quad (1)$$

We obtain this transformation through a two steps calibration. In the first step, we record datasets with the headset poses from VICON

[4] developer.microsoft.com/en-us/windows/mixed-reality/calibration

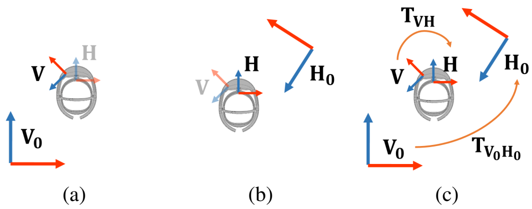


Figure 2: Poses of the $\mathbf{V}$ coordinate system are defined relative to VICON tracking origin $\mathbf{V_0}$ (a), while poses of the $\mathbf{H}$ coordinate system are defined relative to the HoloLens tracking origin $\mathbf{H_0}$ (b). To find the calibration matrix $\mathbf{T_{VH}}$, we need to find the mapping $\mathbf{T_{V_0H_0}}$ from $\mathbf{H_0}$ to $\mathbf{V_0}$ so that $\mathbf{H}$ can be represented in $\mathbf{V_0}$, and $\mathbf{V}$ can be represented in $\mathbf{H_0}$ (c). Now, the virtual camera can be set in $\mathbf{V_0}$ with $\mathbf{T_{V_0H}} = \mathbf{T_{V_0V}} \cdot \mathbf{T_{VH}}$.

and the HoloLens tracking. The user mounts the HoloLens and performs asymmetric translations along the three axis of movement while rotations are avoided. The poses of the two tracked objects are represented in distinct coordinate systems, $\mathbf{V}$ in $\mathbf{V_0}$ (Fig. 2a) and $\mathbf{H}$ in $\mathbf{H_0}$ (Fig. 2b). To align their axis of movement (Fig. 2c), we need to estimate the rotation $\mathbf{R_{V_0H_0}}$ in $\mathbf{p}'_{v_i} = \mathbf{c}_h + \mathbf{R_{V_0H_0}} \cdot (\mathbf{p}_{v_i} - \mathbf{c}_v)$ that minimizes $E = \sum_{i=1}^{n} (\mathbf{p}_{h_i} - \mathbf{p}'_{v_i})^2$, where $\mathbf{p}_v$ and $\mathbf{p}_h$ are positions of $\mathbf{V}$ and $\mathbf{H}$ in $\mathbf{V_0}$ and $\mathbf{H_0}$, and $\mathbf{c}_v$ and $\mathbf{c}_h$ are their respective centroids.

The matrix $\mathbf{R_{V_0H_0}}$ is obtained with the singular value decomposition (SVD) of the 3×3 covariance matrix $\mathbf{M} = \mathbf{P}_v \cdot \mathbf{P}_h^T$, where $\mathbf{P}_v$ and $\mathbf{P}_h$ are $3 \times n$ matrices whose the i-th column is obtained with $\mathbf{p}_{v_i} - \mathbf{c}_v$ and $\mathbf{p}_{h_i} - \mathbf{c}_h$, respectively. SVD gives $\mathbf{M} = \mathbf{U} \cdot \Sigma \cdot \mathbf{V}^T$, where the optimal – in terms of minimizing the sum of squared errors – $\mathbf{R_{V_0H_0}}$ is given by $\mathbf{R_{V_0H_0}} = \mathbf{V} \cdot \mathbf{U}^T$ [1].

Consequently, the VICON and HoloLens trajectories can be aligned in terms of movement direction using the $\mathbf{R_{V_0H_0}}$ matrix. This allows us to estimate a linear mapping $\mathbf{R}_i$ from the $\mathbf{V}$ to $\mathbf{H}$ for each corresponding point on the aligned trajectories with $\mathbf{R}_i = \mathbf{R}_{v_i}^T \cdot \mathbf{R_{V_0H_0}}^T \cdot \mathbf{R}_{h_i}$, where $\mathbf{R}_{v_i}$ and $\mathbf{R}_{h_i}$ refer to the $\mathbf{V}$ and $\mathbf{H}$ rotations at corresponding frames. Outlier rotations are removed based on the deviation of angle $\theta_i = arccos(\frac{trace(\mathbf{R}_i)-1}{2})$. We define inlier rotations as $Q_1 - IQR * 1.5 \leq inliers \leq Q_3 + IQR * 1.5$, where Q_1 and Q_3 are the first and third quartiles of the set of angles, and $IQR = Q_3 - Q_1$. The remaining rotations are converted into the quaternion representation $\mathbf{q}_i$, and their baricentric mean $\overline{\mathbf{q}}$ is used to approximate the average rotation $\mathbf{R}$.

In the second step, the headset is rigidly attached to a tripod, which allows for spherical movements around a spherical joint. We fit a sphere to each collection of positions $\mathbf{p}_v$ and $\mathbf{p}_h$ (positions of $\mathbf{V}$ and $\mathbf{H}$ coordinate systems). By subtracting the center of rotation from the positions, we obtain the $3 \times N$ matrices $\mathbf{P}_v$ and $\mathbf{P}_h$, and each pair of poses can be used to build a $3 \times N$ matrix $\mathbf{M}$, with $\mathbf{M}_i = \mathbf{R}_{v_i}^T \cdot (\mathbf{R_{V_0H_0}}^T \cdot \mathbf{P}_{h_i} - \mathbf{P}_{v_i})$ for each i-th column of $\mathbf{M}$. Finally, the mean value of the rows of $\mathbf{M}$ approximates the translation $\mathbf{t}$ that defines the physical center of projection of the headset relative to the VICON arbitrary tracker depicted in Fig. 1. As a result, we can build the rigid transformation $\mathbf{T_{VH}}$ in Equation 1. Note that this procedure depends on both HoloLens and VICON tracking, and is therefore susceptible to HoloLens tracking limitations when a small amount of data is used, improved transformations can be obtained with the mean of several $\mathbf{V}$ and $\mathbf{H}$ pose correspondences in multiple recordings. Fig. 3 shows the calibration in use for a visualization application (detailed in [3]). Recordings were made by frame ($\approx 60Hz$), with the latest value of both tracking systems. As VICON tracking presents a higher latency, latency estimation was performed with the cross-correlation of the speed of pairs of recordings.

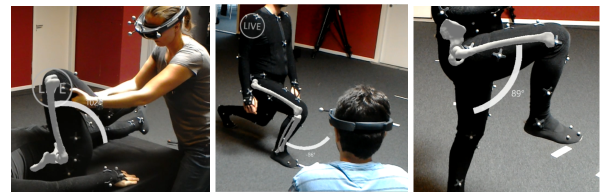


Figure 3: Sample of results in an AR visualization application.

3 Calibration Quality

The built-in camera of the HoloLens headset was used to estimate the projection error of the calibration procedure. The HoloLens overlays the camera image with the virtual scene as rendered for the right eye, capturing an image of the real environment that is superimposed by the virtual information. The HoloLens and a reference object were placed at the two ends of a 70 cm long rigid structure, both tracked using the VICON system. The reference object was placed so that a spherical marker (⌀ 14 mm) on its center could be seen close to the center of the built-in HoloLens camera image. A virtual representation of the marker was rendered in the virtual space based on the tracked position of the reference object. The quality of calibration was assessed based on the overlap of the real marker with its virtual counterpart.

A total of 12 frames were randomly selected from a video with the rigid structure at 12 different orientations while at the center of the VICON tracking volume. Our results showed a mean absolute projection error of $3mm$ ($SD = 1mm$) between the real and virtual spheres.

4 Conclusion

We presented a calibration procedure to register consumer HMDs to 3rd party tracking systems. We further demonstrated its effectiveness with a HoloLens, allowing to accurately overlay real tracked objects with a rendered virtual representation (error $\approx$ 3 mm when 70 cm away from the HoloLens). We note that as the calibration relies on the tracking quality of the both, built-in and 3rd party systems, the calibration is subject to the tracking quality. Poor tracking quality can be mitigated by using the mean transformation of multiple tracking datasets and calibrations. We note that further quality experiments should be performed to define clearer guidelines and to optimize the selection of the calibration on such situations.

Acknowledgments

This work was supported by La Tour Hospital and the clinics of Hirschlanden La Colline.

References

[1] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(5):698–700, Sept 1987.

[2] J. C. Chou and M. Kamel. Quaternions approach to solve the kinematic equation of rotation of a sensor-mounted robotic manipulator. In *Robotics and Automation, 1988. Proceedings., 1988 IEEE International Conference on*, pp. 656–662. IEEE, 1988.

[3] H. G. Debarba, M. E. de Oliveira, A. Lädermann, S. Chagué, and C. Charbonnier. Augmented reality visualization of joint movements for physical examination and rehabilitation. In *2018 IEEE Virtual Reality (VR)*. IEEE, 2018.

[4] L. Qian, E. Azimi, P. Kazanzides, and N. Navab. Comprehensive tracker based display calibration for holographic optical see-through head-mounted display. *arXiv preprint arXiv:1703.05834*, 2017.

[5] M. Tuceryan, Y. Genc, and N. Navab. Single-point active alignment method (spaam) for optical see-through hmd calibration for augmented reality. *Presence: Teleoperators and Virtual Environments*, 11(3):259–276, 2002.