Dataset: HoloSet - A Dataset for Visual-Inertial Pose Estimation in Extended Reality

Yasra Chandio
University of Massachusetts, Amherst
ychandio@umass.edu

Noman Bashir
University of Massachusetts, Amherst
nbashir@umass.edu

Fatima M. Anwar
University of Massachusetts, Amherst
fanwar@umass.edu

ABSTRACT

There is a lack of datasets for visual-inertial odometry applications in Extended Reality (XR). To the best of our knowledge, there is no dataset available that is captured from an XR headset with a human as a carrier. To bridge this gap, we present a novel pose estimation dataset — called `HoloSet` — collected using Microsoft Hololens 2, which is a state-of-the-art head mounted device for XR. Potential applications for `HoloSet` include visual-inertial odometry, simultaneous localization and mapping (SLAM), and additional applications in XR that leverage visual-inertial data.

`HoloSet` captures both macro and micro movements. For macro movements, the dataset consists of more than 66,000 samples of visual, inertial, and depth camera data in a variety of environments (indoor, outdoor) and scene setups (trails, suburbs, downtown) under multiple user action scenarios (walk, jog). For micro movements, the dataset consists of more than 12,000 samples of additional articulated hand depth camera images while a user plays games that exercise fine motor skills and hand-eye coordination. We present basic visualizations and high-level statistics of the data and outline the potential research use cases for `HoloSet`.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; • **Information systems → Data mining**; **Spatial-temporal systems**; • **General and reference** → *Measurement*.

KEYWORDS

Dataset, Extended Reality, AR/VR, Tracking, SLAM, Odometry, Computer Vision, Navigation, Deep Learning, Hololens.

ACM Reference Format:
Yasra Chandio, Noman Bashir, and Fatima M. Anwar. 2022. Dataset: HoloSet - A Dataset for Visual-Inertial Pose Estimation in Extended Reality. In *The 20th ACM Conference on Embedded Networked Sensor Systems (SenSys '22), November 6–9, 2022, Boston, MA, USA*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3560905.3567763

1 INTRODUCTION

Emerging technologies such as Extended Reality (XR) leverage macro- and micro-level tracking to enable useful applications in healthcare [1, 31, 52], manufacturing [20, 27, 36], education [3, 24, 32], and gaming [12, 41]. Tracking algorithms used in such applications rely on multidimensional spatio-temporal relationships among multimodal sensing streams from a combination of visual cameras and inertial sensors. State-of-the-art tracking algorithms typically use machine learning or deep learning algorithms that require a lot of data for training [10, 43]. However, despite the significant scientific and commercial interest in XR applications, there is a lack of public datasets that facilitate XR research and allow the development of novel applications.

There are several visual and inertial datasets available for the related fields of autonomous vehicles and mobile robotic systems, such as KITTI [19], ADVIO [15], OxIOD [11], YTU [21], and TUM VI [42]. These datasets primarily provide camera images and Inertial Measurement Unit (IMU) data collected using hand-held devices or vehicles. These datasets mainly target visual-inertial odometry (VIO) [10, 14] and simultaneous localization and mapping (SLAM) [2, 40]. However, this data is not suited for XR applications where a user wears a Head-mounted Device (HMD) and requires an understanding of the surrounding environment. Furthermore, these datasets lack depth images, hand and eye tracking data, which is required for spatial mapping, scene understanding, and keylogging tasks. The data and capabilities mentioned above are essential to enable novel, interactive, and useful applications in XR.

To bridge this gap, we collect and publicly release a dataset, `HoloSet`, using Microsoft Hololens 2. Holoset is designed and collected for academic and industrial researchers exploring new ideas in the fields of VIO and SLAM, and other computer vision techniques used in XR. We collect sensor data using a state-of-the-art HMD - Hololens 2 [25] by capturing the raw synchronized data streams from the following sensors: a depth camera, an RGB camera, four grayscale visible light tracking (VLC) cameras, and an IMU consisting of an accelerometer, gyroscope, and magnetometer. Additionally, we record the ground truth pose trajectory. In the future, we plan to add hand-tracking and eye-tracking data. For macro movements, we provide more than 66,000 samples of data in a variety of environments (indoor, outdoor) and scene setups (woods, suburbs, downtown) under multiple user locomotion scenarios (walk, jog). For micro-movements, we provide more than 12,000 samples of data while a user plays games that exercise fine motor skills and hand-eye coordination. In collecting and releasing our dataset, we make the following contributions.

First contribution. A large dataset captured at a high frame rate using an IMU sensor, one RGB camera, four grayscale cameras, and a depth camera. To the best of our knowledge, this is the first dataset collected using a head mounted device. Our dataset is available at Zenodo [17], under the DOI 10.5281/zenodo.7200131[1].

SenSys '22, November 6–9, 2022, Boston, MA, USA

ACM ISBN 978-1-4503-9886-2/22/11…$15.00
https://doi.org/10.1145/3560905.3567763

[1]https://tinyurl.com/holoset-dataset

	KITTI [19]	ADVIO [15]	OxIOD [11]	HoloSet (*this paper*)
Year	2012	2018	2018	2022
Carrier	Car	Hand-held	Hand-held, bag, pocket, trolley	Head-mounted device
Environment	Indoors/Outdoors	Indoors/Outdoors	Indoors	Indoors/Outdoors
Movement type	Macro	Macro	Macro	Macro/Micro
User actions	Drive	walk	Halt, walk, jog, run	Walk, jog + play games that exercise fine motor skills and hand-eye coordination
Scene setup	City-scale	Multiple buildings + outdoor scenes	Office buildings	Multiple levels in 4 buildings + city center, urban scenes, and hiking trails + micro movements inside a room
Hardware setup	Custom [19]	Smartphones	Smartphones	Microsoft Hololens 2
Data types	Camera images, laser scans, point cloud, IMU and GPS data	Camera images, IMU data, barometer data	IMU data	RGB, depth, and gray-scale (4×) camera images, IMU data
Total distance	∼24 miles	∼2.8 miles	26+ miles	4miles
Ground truth	GPS/IMU	IMU + position fixes	Vicon	Hololens pose

Table 1: *An overview of – and comparison with – the related datasets.*

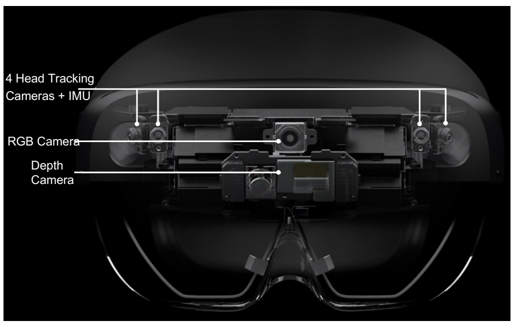


Figure 1: Cameras and IMU sensor position on the HoloLens 2 headset (figure source [25]).

Second contribution. To the best of our knowledge, we provide the first dataset that captures micro-movements. We provide more than 12,000 samples of data where a user plays Jenga and Operation games that exercise fine motor skills.
Third contribution We provide a post-processing script that leverages data conditioning techniques, i.e., synchronizing raw sensor data across sensing modalities, to the raw data and provides clean synchronized version of the data.
Fourth contribution. We outline a comprehensive list of future applications where `HoloSet` can enable extended reality (and other mobile and wearable device-based) applications.

2 RELATED WORK

There are several datasets that provide visual and inertial data to assist research in XR, mobile devices, and other wearable devices. Some dastasets are used to develop visual-inertial odometry (VIO) and SLAM algorithms such as KITTI [19], ADVIO [15], OxIOD [11], YTU [21], and TUM VI [42], Oxford RobotCar [33], EuRoC MAV [7], UMA-VI [57], PALVIO [54], ICL-NUIM [22], and Málaga [4]. Other include datasets that focus on human gait (MAREA [28], OU-ISIR [49]), occupancy (LARA [39]), and activity recognition (USC-HAD [55], CMU-MMAC [16], Opportunity [9]).

However, in this section, we only discuss three closely related VIO datasets (KITTI [19], ADVIO [15], OxIOD [11]). We briefly summarize and compare these datasets to *HoloSet* (our dataset) in Table 1. KITTI is a state-of-the-art benchmark dataset with data collected in both indoor and outdoor environments. However, its sensors are rigidly fixed to the chassis, that makes it suitable for studying vehicle movements, but not directly applicable to studying human movements (like ours). ADVIO and OxIOD datasets are collected using the handheld devices that make them suitable for human movement research. However, ADVIO data only provides pseudo ground truth generated by their state estimation algorihtm that only used inertial odometry. OxIOD provides highly accurate ground truth generated using Vicon [35], but it only has inertial data. In addition, OxIOD and ADVIO provide highly processed handheld smartphone data that may hide actual nuances in readings. More importantly, they are not collected using a head-mounted display.

In `HoloSet`, we collect data using a head-mounted device (Hololens 2) that offers raw sensor data from 6 cameras (covering multiple views), an IMU, and highly accurate ground truth (2-4 cm error [23, 45]). `HoloSet` also has diverse macro movements (walk normally, slowly or jog, as well as halting) and wide range of scenario covering various indoor and outdoor environments. `HoloSet` also includes micro movement data that have articulated hand movement with depth cameras. It also offers a large number of samples, 80,000+, making it suitable for deep learning approaches, which require large amounts of data and high-accuracy labels. In summary, `HoloSet` better represents human motion in everyday situations and provides a large number of samples to enable a wide range of applications that may use simple to complex models.

3 HOLOSET DATA COLLECTION SETUP

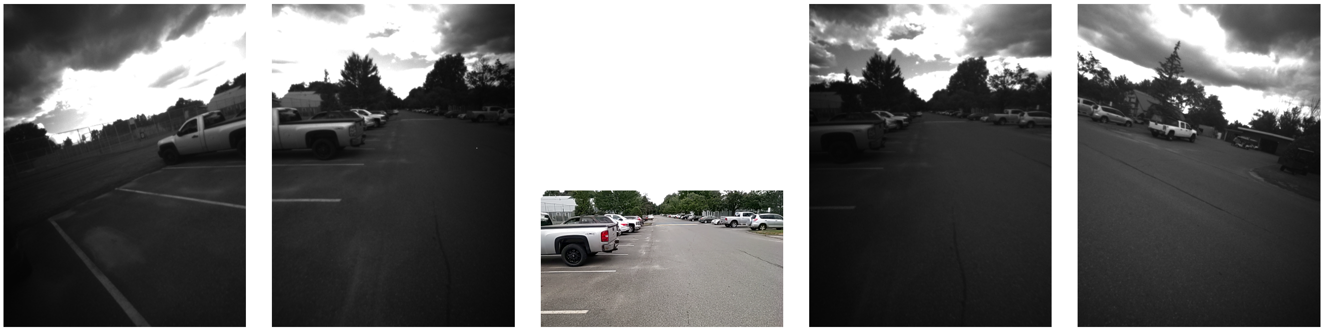
In this section, we provide the data collection setup and detail the data conditioning details for the `HoloSet`.

3.1 Data collection setup

We collect data from HoloLens 2 headset in research mode mounted on the head of a user. Figure 1 shows the position of the various cameras and sensors on the headset. We next describe the types of data that we collect from headset and the software setup used to collect the data across different settings, user actions, and environments.

3.1.1 ***Visual.*** Our visual data comes from two types of cameras. First, we collect data from four visible light-tracking cameras (VLC) that produce grayscale images at 30 frames per second. The second source of visual data is HoloLens's Photo-to-Video (PV) RGB stream which is generated using an 8Mpix RGB camera. We capture the images from the stream at 30-45 frames per second.

3.1.2 ***Inertial.*** The inertial data is collected from the HoloLens's Inertial Measurement Unit (IMU), which reports data from accelerometer, gyroscope, and magnetometer.



(a) left left (b) left front (c) photo video (d) right front (e) right right

Figure 2: Samples images from a suburban walk sequences for the four VLC grey-scale cameras – (a) left left, (b) left front, (d) right front, (e) right right — and photo video RGB camera, (c) photo video.

3.1.3 ***Depth****.* HoloLens 2 has a 1Mpix depth camera that can operate in two modes: Articulated HAnd Tracking (AHAT) and Long Throw. AHAT mode provides a near-depth sensing images at 45 frames per second. Long Throw mode provides far-depth sensing images at 1-5 frames per second rate. HoloLens uses AHAT mode for hand-tracking and Long Throw mode to compute spatial mapping. The depth camera can operate in only one of the modes at a given time. We collect depth camera images in Long Throw mode when capturing macro movements and in AHAT mode when capturing micro movements.

3.1.4 ***Ground truth****.* We report the pose values reported by the HoloLens 2 as the ground-truth for our data. There is a lot of prior work that evaluates the accuracy of the HoloLens 1 and HoloLens 2 tracking algorithm and shows that the reported position data is highly accurate with an error in the range of 2-4 cm [23, 45]. This accuracy is highly sufficient for all potential tracking applications.

Further details about the HoloLens's hardware setup and the data it provides can be found in the HoloLens research mode paper [50].

3.2 Data conditioning

In this section, we detail the data conditioning script that we provide alongside the raw version of the data to generate the synchronized version of `HoloSet`.

3.2.1 ***Cleaning.*** We observe that the IMU sensor of `HoloLens` 2 can erroneously generate abnormally large values (>10^5). If we observe more than five consecutive abnormal values or more than 5% abnormal samples in the whole sequence, we discard the data and regenerate the entire sequence. For other cases, we replace the missing values using linear interpolation.

3.2.2 ***Synchronization.*** Our data collection setup collects data from different sensors and cameras at different rates. The PV RGB stream has the highest frame rate at $\sim$ 30 images per second, and we use its timestamps as a reference to synchronize the data. For IMU, VLC cameras, and depth cameras, we simply find the sensor measurements and images closer to the particular PV timestamp. The lowest frame rate in our dataset was ~20 samples per second for the magnetometer, which means that the maximum time difference between the PV data and any other sensor can be 2.5 ms. The synchronized data is useful for training machine learning models which often require different input features to have equally-spaced and equal number of samples. We also provide the raw data and the matched timestamps in `HoloSet`.

3.2.3 ***Data privacy.*** We anonymize the faces of people in the images using Python's OpenCV library [5]. We use Haar feature-based cascade classifiers to detect faces in the images [44] and blur them using OpenCV's `blur` function for PV and VLC cameras. The data set does not contain identifiable human information.

4 HOLOSET DATASET

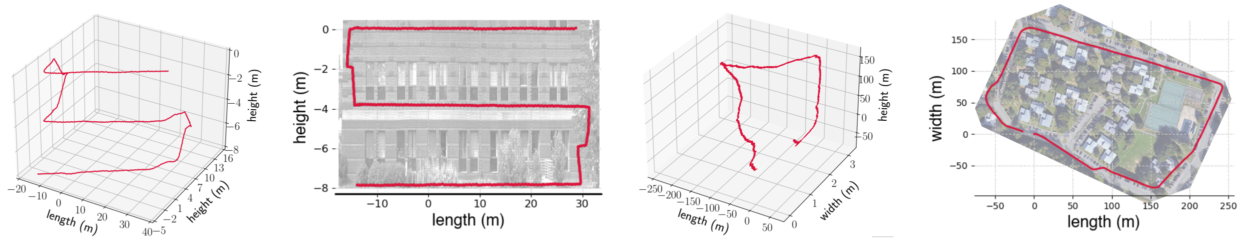
In this section, we provide an overview of the different settings in which we collect data, report key statistics on the dataset, and provide sample visualizations of the collected data.

4.1 Overview

The goal of our dataset is to capture micro and macro movements of a user while wearing an Extended Reality (XR) Head Mounted Device (HMD) such as Microsoft HoloLens 2 [25]. The macro and micro movements result from, and are of interest to, different kinds of applications. Therefore, we collect the data in two different setups that emphasize the respective movements. Tables 1 & 2 describes the environment, the scene setup and the user actions for `HoloSet`.

4.1.1 ***Macro movements.*** These movements refer to an activity that moves the entire body of a user from one physical location to another, also termed *locomotion*. The most common examples of *locomotion* are walking, jogging, and running (*User Actions* in Table 1). Locomotion can occur indoors or outdoors. For the outdoors, we consider three types of scene: a natural setting of a hiking trail, a suburban street with moderate human and vehicular traffic, and finally the city center of a university town. These three scene setups have been the focus of all the previous datasets and cover most of the current and future XR applications. For indoor settings, we consider three environments: a low human-traffic environment of a research building, a high human-traffic environment of a campus center, and moderate human-traffic of a student center.

4.1.2 ***Micro movements.*** These movements refer to the fine movement of a specific part of the user's body, such as hands and eyes. Inertial data has a strong temporal component, but it lacks



(a) research building (3D view) (b) research building (side view) (c) suburban walk (3D view) suburban walk (top view)

Figure 3: Ground-truth trajectories of sequences collected while walking down from the third to the ground floors inside a 3-floor research building, (a) 3D view and (b) side view, and walking anti-clockwise inside a suburban housing society, (c) 3D view and (d) top view.

Environment	Action	Seqs	Samples	Distance (meters)
hiking trail	walk	3	8518	686
suburbs	walk	3	15150	1540
	jog	5	8251	1003
city center	walk	3	5132	407
outdoors	-	**14**	**37051**	**3636**
campus center	walk	2	5843	497
student center	walk	2	5664	486
research center	walk	4	17419	1648
indoors	-	**8**	**28926**	**2631**
room	jenga	5	10842	-
	operation	2	1726	-
micro	-	**7**	**12568**	-
all	**all**	**29**	**78545**	**6267**

Table 2: `HoloSet` ***key characteristics.***

the ability to perceive subtle spatial changes that occur due to micro movements, which are of interest to applications where users engage in complex and intricate actions, such as surgery. To mimic the behavior of a user in such applications, our user plays two board games that exercise a user's hand-eye coordination and fine motor skills. The games that the user plays are Classic Jenga [58] and Operation [46]. The user plays these games multiple times on a table in a room. Currently, we provide depth images with articulated hand movement tracking captured through a depth camera [50]. In the future, we will add hand and eye tracking data to `HoloSet`.

4.2 Key statistics

Table 2 provides the high-level statistics of the collected dataset. It contains 9 different combinations of environment, action, and movement types. We collected a total of 14 sequences in outdoor settings that contained more than 37k samples collected over the course of 3636 meters. The indoor dataset contains 8 sequences consisting of almost 29k samples collected over 2631m distance inside three buildings. For micro movements, we collected 7 sequences consisting of 12.5k samples. Overall, our dataset contains 29 sequence and 78.5k samples that cover more than 6200 meters.

4.3 Sample visualizations

Figure 2 shows sample images from the 4 VLC (grey-scale) and 1 PV (RGB) cameras in a single instance of the suburban walk sequence. The stereo and peripheral setting of the VLC cameras captures the complete view of the user's path, as opposed to a single RGB image typically provided by most datasets. HoloLens 2 also generates an RGB image that captures the scene in front of the user.

Figure 3 shows sample ground-truth 3D and 2D trajectories from an indoor walk in a research building and an outdoor walk in a suburban housing society. The end-to-end length of the research building is ~55m (~180ft) and has a height of ~13m (~430ft). The 3D trajectory Figure 3(a) shows the movement along the length of the research building and towards the stairs. Since this walk started at the top floor, the sideways view Figure 3(b) along the xz-plane shows the drop in height over the course of the walk (-ve y-axis).

Similarly, the 3D trajectory Figure 3(c) for the outdoor walk shows the change in height over the course of the walk due to the hilly nature of the area. The top view Figure 3(d) shows the user trajectory along the xy-plane. The ground-truth captures the fine changes in the user trajectory over the course of the walk.

4.4 Future work

Our current version of the dataset contains sequences from a wide-range of environments, as listed in Table 2. However, we plan to keep updating `HoloSet` in multiple ways. First, we plan to add more sequences for the same environment and user settings. We will also add data for the same user under different settings and actions such as walking on an indoor walk/jog track inside the campus recreation center of our university. Second, we will add data for more users for all environment and action combinations. Third, we will welcome suggestions from the research community for new user, environment, and action settings. We will satisfy reasonable data requests at our earliest convenience. Finally, we will add hand- and eye-tracking data in the future to enable a wide-ranging new set of applications.

5 POTENTIAL USE CASES

Current tasks of interest for this dataset are: visual inertial odometry, SLAM and tracking. Our datasets' outdoor macromovement sequences could be useful for several applications including pedestrian heading estimation [53], human velocity recognition [18], outdoor navigation system [6], 3d tracking and forecasting [8], and walking direction estimation [34]. Possible usecase for indoor macro movement sequences are: pose estimation for ground robots[56], human location recognition [29], indoor positioning and navigation in environments like residential buildings, offices [26, 51] Our micro movement scenarios can be useful in developing tools for surgical planning, training and real-time procedure [52], remote collaboration on factory floor and warehouse [37, 47], and remote maintenance [38].

In summary Holoset can provide benefit in advancing overall research and development for real-world tools that may or not belong to XR. Various such field of interest are stereo visual odometry based on motion [13], deep learning based visual odometry [30], feature-based visual odometry [10], visual SLAM for monocular, stereo, and RGB cameras [48], and sensor fusion [14]. All of these technique require diverse and large amount of data to be trained to sufficient accuracy, that HoloSet fills that gap. Furthermore, complex deep learning model can benefit from the multiple views of the camera images that `HoloSet` provides.

6 CONCLUSION

In this paper, we presented `HoloSet`, a large dataset captured at a high frame rate using an IMU sensor, one RGB camera, four grayscale cameras, and a depth camera. To the best of our knowledge, this is the first dataset collected using a head mounted device. An additional novelty of the dataset also lies in capturing micro-movements while a user plays Jenga and Operation games that exercise fine motor skills. To facilitate the use of data, we provide a post-processing script that leverages data conditioning techniques, i.e., synchronizing raw sensor data across sensing modalities, to the raw data and provides clean synchronized version of the data. We inspire future research in this domain, we outline a comprehensive list of future applications where `HoloSet` can enable extended reality (and other mobile and wearable device-based) applications. We currently release the raw data and conditioning code, which is available at Zenodo under the DOI 10.5281/zenodo.7200131. We will plan to provide an active support to `HoloSet` users and add more scenarios to the existing dataset.

Acknowledgements. This research is supported by NSF grant 2230143.

REFERENCES

[1] 2019. Real life Dangers of Augmented Reality. https://spectrum.ieee.org/consumer-electronics/portable-devices/the-reallife-dangers-of-augmented-reality

[2] Tim Bailey and Hugh Durrant-Whyte. 2006. Simultaneous localization and mapping (SLAM): Part II. *IEEE robotics & automation magazine* 13, 3 (2006), 108–117.

[3] James Birt, Zane Stromberga, Michael Cowling, and Christian Moro. 2018. Mobile mixed reality for experiential learning and simulation in medical and health sciences education. *Information* 9, 2 (2018), 31.

[4] José-Luis Blanco-Claraco, Francisco-Angel Moreno-Duenas, and Javier González-Jiménez. 2014. The Málaga urban dataset: High-rate stereo and LiDAR in a realistic urban scenario. *The International Journal of Robotics Research* 33, 2 (2014), 207–214.

[5] Gary Bradski. 2000. The openCV library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer* 25, 11 (2000), 120–123.

[6] Martin Brossard, Axel Barrau, and Silvere Bonnabel. 2019. RINS-W: Robust inertial navigation system on wheels. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2068–2075.

[7] Michael Burri, Janosch Nikolic, Pascal Gohl, Thomas Schneider, Joern Rehder, Sammy Omari, Markus W Achtelik, and Roland Siegwart. 2016. The EuRoC micro aerial vehicle datasets. *The International Journal of Robotics Research* 35, 10 (2016), 1157–1163.

[8] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, et al. 2019. Argoverse: 3d tracking and forecasting with rich maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8748–8757.

[9] Ricardo Chavarriaga, Hesam Sagha, Alberto Calatroni, Sundara Tejaswi Digumarti, Gerhard Tröster, José del R Millán, and Daniel Roggen. 2013. The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters* 34, 15 (2013), 2033–2042.

[10] Changhao Chen, Stefano Rosa, Yishu Miao, Chris Xiaoxuan Lu, Wei Wu, Andrew Markham, and Niki Trigoni. 2019. Selective sensor fusion for neural visual-inertial odometry. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10542–10551.

[11] Changhao Chen, Peijun Zhao, Chris Xiaoxuan Lu, Wei Wang, Andrew Markham, and Niki Trigoni. 2018. OxIOD: The Dataset for Deep Inertial Odometry. *arXiv preprint arXiv:1809.07491* (2018).

[12] Adrian David Cheok, Anuroop Sreekumar, Cao Lei, and Le Nam Thang. 2006. Capture the flag: mixed-reality social gaming with smart phones. *IEEE Pervasive Computing* 5, 2 (2006), 62–69.

[13] Wenyan Ci, Yingping Huang, and Xing Hu. 2019. Stereo visual odometry based on motion decoupling and special feature screening for navigation of autonomous vehicles. *IEEE Sensors Journal* 19, 18 (2019), 8047–8056.

[14] Ronald Clark, Sen Wang, Hongkai Wen, A. Markham, and Agathoniki Trigoni. 2017. VINet: Visual-Inertial Odometry as a Sequence-to-Sequence Learning Problem. In *AAAI*.

[15] Santiago Cortés, Arno Solin, Esa Rahtu, and Juho Kannala. 2018. ADVIO: An Authentic Dataset for Visual-Inertial Odometry. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

[16] Fernando De la Torre, Jessica Hodgins, Adam Bargteil, Xavier Martin, Justin Macey, Alex Collado, and Pep Beltran. 2009. Guide to the carnegie mellon university multimodal activity (cmu-mmac) database. (2009).

[17] European Organization For Nuclear Research and OpenAIRE. 2013. Zenodo. https://doi.org/10.25495/7GXK-RD71

[18] Tobias Feigl, Sebastian Kram, Philipp Woller, Ramiz H Siddiqui, Michael Philippsen, and Christopher Mutschler. 2020. RNN-aided human velocity estimation from a single IMU. *Sensors* 20, 13 (2020), 3656.

[19] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. 2013. Vision Meets Robotics: The KITTI Dataset. *The International Journal of Robotics Research* (2013).

[20] Mar Gonzalez-Franco, Rodrigo Pizarro, Julio Cermeron, Katie Li, Jacob Thorn, Windo Hutabarat, Ashutosh Tiwari, and Pablo Bermell-Garcia. 2017. Immersive mixed reality for manufacturing training. *Frontiers in Robotics and AI* 4 (2017), 3.

[21] Mert Gurturk, Abdullah Yusefi, Muhammet Fatih Aslan, Metin Soycan, Akif Durdu, and Andrea Masiero. 2021. The YTU Dataset and Recurrent Neural Network based Visual-inertial Odometry. *Measurement* (2021).

[22] Ankur Handa, Thomas Whelan, John McDonald, and Andrew J Davison. 2014. A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM. In *2014 IEEE international conference on Robotics and automation (ICRA)*. IEEE, 1524–1531.

[23] Patrick Hübner, Kate Clintworth, Qingyi Liu, Martin Weinmann, and Sven Wursthorn. 2020. Evaluation of HoloLens tracking and depth sensing for indoor mapping applications. *Sensors* 20, 4 (2020), 1021.

[24] Charles E Hughes, Christopher B Stapleton, Darin E Hughes, and Eileen M Smith. 2005. Mixed reality in education, entertainment, and training. *IEEE computer graphics and applications* 25, 6 (2005), 24–30.

[25] Microsoft Inc. 2020. Hololens 2. https://www.microsoft.com/en-us/hololens/hardware. (2020). "[Online; accessed 13th-June-2022]".

[26] SJ Ingram, D Harmer, and M Quinlan. 2004. Ultrawideband indoor positioning systems and their use in emergencies. In *PLANS 2004. Position Location and Navigation Symposium (IEEE Cat. No. 04CH37556)*. IEEE, 706–715.

[27] Max Juraschek, Lennart Büth, Gerrit Posselt, and Christoph Herrmann. 2018. Mixed reality in learning factories. *Procedia Manufacturing* 23 (2018), 153–158.

[28] Siddhartha Khandelwal and Nicholas Wickström. 2017. Evaluation of the performance of accelerometer-based gait event detection algorithms in different real-world scenarios using the MAREA gait database. *Gait & posture* 51 (2017), 84–90.

[29] Itzik Klein. 2019. Smartphone location recognition: A deep learning-based approach. *Sensors* 20, 1 (2019), 214.

[30] Nicola Krombach, David Droeschel, Sebastian Houben, and Sven Behnke. 2018. Feature-based visual odometry prior for real-time semi-dense stereo SLAM. *Robotics and Autonomous Systems* 109 (2018), 38–58.

[31] Cristian A Linte, Katherine P Davenport, Kevin Cleary, Craig Peters, Kirby G Vosburgh, Nassir Navab, Philip Eddie Edwards, Pierre Jannin, Terry M Peters, David R Holmes, and Richard A Robb. 2013. On Mixed Reality Environments for Minimally Invasive Therapy Guidance: Systems Architecture, Successes and Challenges In Their Implementation From Laboratory To Clinic. *Computerized Medical Imaging and Graphics* 37, 2 (March 2013), 83–97. https://doi.org/10.1016/j.compmedimag.2012.12.002

[32] Melanie J Maas and Janette M Hughes. 2020. Virtual, augmented and mixed reality in K–12 education: A review of the literature. *Technology, Pedagogy and Education* 29, 2 (2020), 231–249.

[33] Will Maddern, Geoffrey Pascoe, Chris Linegar, and Paul Newman. 2017. 1 year, 1000 km: The Oxford RobotCar dataset. *The International Journal of Robotics Research* 36, 1 (2017), 3–15.

[34] Adi Manos, Tamir Hazan, and Itzik Klein. 2022. Walking Direction Estimation Using Smartphone Sensors: A Deep Network-Based Framework. *IEEE Transactions on Instrumentation and Measurement* 71 (2022), 1–12.

[35] Pierre Merriaux, Yohan Dupuis, Rémi Boutteau, Pascal Vasseur, and Xavier Savatier. 2017. A study of vicon system positioning performance. *Sensors* 17, 7 (2017), 1591.

[36] Dimitris Mourtzis, John Angelopoulos, and Nikos Panopoulos. 2021. Collaborative manufacturing design: a mixed reality and cloud-based framework for part design. *Procedia CIRP* 100 (2021), 97–102.

[37] Dimitris Mourtzis, Aikaterini Vlachou, and Vasilios Zogopoulos. 2017. Cloud-based augmented reality remote maintenance through shop-floor monitoring: a product-service system approach. *Journal of Manufacturing Science and Engineering* 139, 6 (2017).

[38] Dimitris Mourtzis, Vasilios Zogopoulos, and E Vlachou. 2017. Augmented reality application to support remote maintenance as a service in the robotics industry. *Procedia Cirp* 63 (2017), 46–51.

[39] Friedrich Niemann, Christopher Reining, Fernando Moya Rueda, Nilah Ravi Nair, Janine Anika Steffens, Gernot A Fink, and Michael Ten Hompel. 2020. Lara: Creating a dataset for human activity recognition in logistics using semantic attributes. *Sensors* 20, 15 (2020), 4083.

[40] A Alan B Pritsker. 1984. *Introduction to Simulation and SLAM II.* Halsted Press.

[41] David Robert, Ryan Wistorrt, Jesse Gray, and Cynthia Breazeal. 2010. Exploring mixed reality robot gaming. In *Proceedings of the fifth international conference on tangible, embedded, and embodied interaction*. 125–128.

[42] David Schubert, Thore Goll, Nikolaus Demmel, Vladyslav Usenko, Jörg Stückler, and Daniel Cremers. 2018. The TUM VI Benchmark for Evaluating Visual-inertial Odometry. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE.

[43] Thomas Schöps, Torsten Sattler, and Marc Pollefeys. 2019. BAD SLAM: Bundle Adjusted Direct RGB-D SLAM. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 134–144. https://doi.org/10.1109/CVPR.2019.00022

[44] Ali Sharifara, Mohd Shafry Mohd Rahim, and Yasaman Anisi. 2014. A General Review of Human Face Detection Including a Study of Neural Networks and Haar feature-based cascade Classifier in Face Detection. In *2014 International symposium on biometrics and security technologies (ISBAST)*. IEEE.

[45] Inês Soares, Ricardo B. Sousa, Marcelo Petry, and António Paulo Moreira. 2021. Accuracy and repeatability tests on HoloLens 2 and HTC Vive. *Multimodal Technologies and Interaction* 5, 8 (2021), 47.

[46] John Spinello. 1965. Operation Game. https://hasbrogaming.hasbro.com/en-us/product/classic-operation-game:03B9FF5E-5056-9047-F5FF-D57592C4E2FE.

[47] Anna Syberfeldt, Oscar Danielsson, and Patrik Gustavsson. 2017. Augmented reality smart glasses in the smart factory: Product evaluation guidelines and review of available products. *Ieee Access* 5 (2017), 9118–9130.

[48] Zachary Teed and Jia Deng. 2021. Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras. *Advances in Neural Information Processing Systems* 34 (2021), 16558–16569.

[49] Md Uddin, Thanh Trung Ngo, Yasushi Makihara, Noriko Takemura, Xiang Li, Daigo Muramatsu, Yasushi Yagi, et al. 2018. The ou-isir Large Population Gait Database with Real-life Carried Object and its Performance Evaluation. *IPSJ Transactions on Computer Vision and Applications* 10, 1 (2018), 1–11.

[50] Dorin Ungureanu, Federica Bogo, S. Galliani, Pooja Sama, Xin Duan, Casey Meekhof, Jan Stühmer, Thomas J. Cashman, Bugra Tekin, Johannes L. Schönberger, Pawel Olszta, and Marc Pollefeys. 2020. HoloLens 2 Research Mode as a Tool for Computer Vision Research. *ArXiv* abs/2008.11239 (2020).

[51] Tom Van Haute, Eli De Poorter, Pieter Crombez, Filip Lemic, Vlado Handziski, Niklas Wirström, Adam Wolisz, Thiemo Voigt, and Ingrid Moerman. 2016. Performance analysis of multiple Indoor Positioning Systems in a healthcare environment. *International journal of health geographics* 15, 1 (2016), 1–15.

[52] Petr Vávra, Jan Roman, P. Zonča, Peter Ihnát, M. Němec, Kumar Jayant, Nagy Habib, and Ahmed El-Gendi. 2017. Recent Development of Augmented Reality in Surgery: A Review. *Journal of Healthcare Engineering* 2017 (08 2017), 1–9. https://doi.org/10.1155/2017/4574172

[53] Qu Wang, Haiyong Luo, Langlang Ye, Aidong Men, Fang Zhao, Yan Huang, and Changhai Ou. 2019. Pedestrian heading estimation based on spatial transformer networks and hierarchical LSTM. *IEEE Access* 7 (2019), 162309–162322.

[54] Ze Wang, Kailun Yang, Hao Shi, and Kaiwei Wang. 2022. LF-VIO: A Visual-Inertial-Odometry Framework for Large Field-of-View Cameras with Negative Plane. *arXiv preprint arXiv:2202.12613* (2022).

[55] Mi Zhang and Alexander A Sawchuk. 2012. USC-HAD: A daily activity dataset for ubiquitous activity recognition using wearable sensors. In *Proceedings of the 2012 ACM conference on ubiquitous computing*. 1036–1043.

[56] Mingming Zhang, Xingxing Zuo, Yiming Chen, Yong Liu, and Mingyang Li. 2021. Pose estimation for ground robots: On manifold representation, integration, reparameterization, and optimization. *IEEE Transactions on Robotics* 37, 4 (2021), 1081–1099.

[57] David Zuñiga-Noël, Alberto Jaenal, Ruben Gomez-Ojeda, and Javier Gonzalez-Jimenez. 2020. The uma-vi dataset: Visual–inertial odometry in low-textured and dynamic illumination environments. *The International Journal of Robotics Research* 39, 9 (2020), 1052–1060.

[58] Uri Zwick. 2002. Jenga. In *Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms*. 243–246.