



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Motion Tracklet Oriented 6-DoF Inertial Tracking Using Commodity Smartphones

### Citation for published version:

Li, P & Lu, CX 2021, Motion Tracklet Oriented 6-DoF Inertial Tracking Using Commodity Smartphones. in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. SenSys '21, Association for Computing Machinery, Inc, New York, NY, USA, pp. 542–545, 19th ACM Conference on Embedded Networked Sensor Systems 2021, Coimbra, Portugal, 15/11/21. <https://doi.org/10.1145/3485730.3494116>

### Digital Object Identifier (DOI):

[10.1145/3485730.3494116](https://doi.org/10.1145/3485730.3494116)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Peer reviewed version

### Published In:

Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Dataset: Motion Tracklet Oriented 6-DoF Inertial Tracking Using Commodity Smartphones

Peize Li  
University of Edinburgh  
Edinburgh, UK  
lipeize.ue@gmail.com

Chris Xiaoxuan Lu  
University of Edinburgh  
Edinburgh, UK  
xiaoxuan.lu@ed.ac.uk

## ABSTRACT

Motion tracklets are the basic fragments of the track followed by a moving object and constitute various everyday motion behavior. An accurate estimation of motion tracklets in 3-D space can enable a wide range of applications, ranging from human computer interaction to medical rehabilitation. This paper presents a novel dataset for accurate 6-DoF motion tracklet estimation with the inertial sensors on commodity smartphones. The dataset consists of around 100 minutes of handheld motion with 3 predominant types of motion tracklets and accurate ground truth using the Vicon systems. With the presented dataset, we further benchmarked the trajectory estimation using a lightweight neural odometry model, showcasing how the dataset can be used while providing quantitative performance for downstream tasks. Our dataset, toolkit and source code available at <https://github.com/MAPS-Lab/smartphone-tracking-dataset>.

## CCS CONCEPTS

• **Computing methodologies** → Neural networks; • **Hardware** → Sensor applications and deployments; • **Human-centered computing** → Ubiquitous and mobile computing systems and tools.

## KEYWORDS

neural inertial tracking, people-centric sensing, smartphone

### ACM Reference Format:

Peize Li and Chris Xiaoxuan Lu. 2021. Dataset: Motion Tracklet Oriented 6-DoF Inertial Tracking Using Commodity Smartphones. In *The 19th ACM Conference on Embedded Networked Sensor Systems (SenSys '21)*, November 15–17, 2021, Coimbra, Portugal. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3485730.3494116>

## 1 INTRODUCTION

Inertial tracking uses accelerometers and gyroscopes to calculate the position and orientation changes of an object relative to a reference point. While it was originally proposed for high-precision military purposes, today's inertial tracking systems have been able to leverage the low-cost MEMS sensors embedded on wearable

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

ACM DATA Workshop co-located with SenSys'21 and BuildSys'21, November 15–17, 2021, Coimbra, Portugal

© 2021 Association for Computing Machinery.  
ACM ISBN 978-1-4503-9097-2/21/11...\$15.00  
<https://doi.org/10.1145/3485730.3494116>

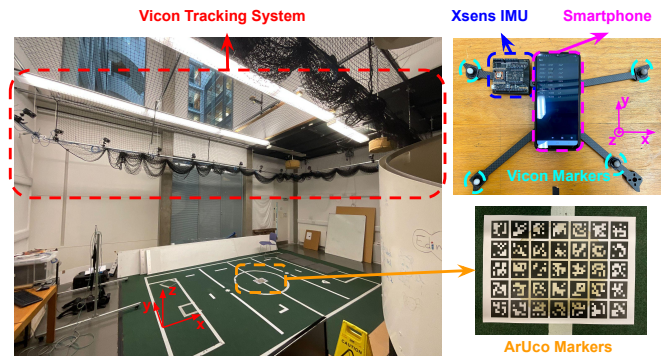
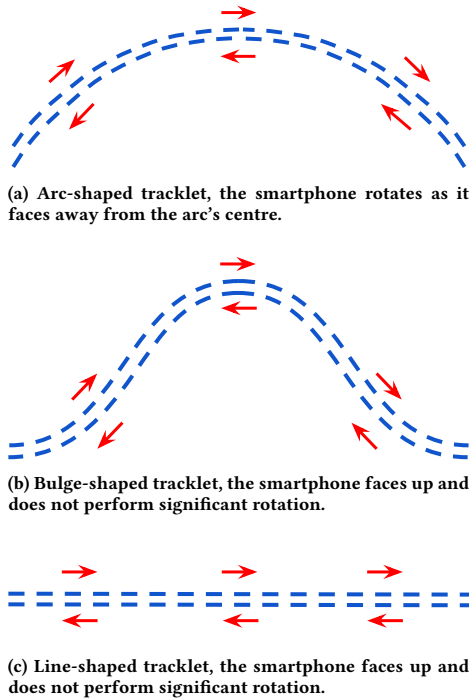


Figure 1: Data recording setup. We choose a mid-range Android smartphone, the Samsung Galaxy A31, as our inertial data collector. Vicon tracking system provides millimeter-level-accuracy ground truth trajectory. Xsens IMU provides high-quality inertial data for comparison with smartphone data. ArUco marker provides secondary ground truth that allows dataset users to expand the dataset.

devices and find a spectrum of applications in everyday world-ranging from medical rehabilitation to human-machine interaction. Among these wearable devices, inertial tracking with commodity smartphones receives the most attention in industry and academia due to their prominent ubiquity and portability.

Unlike the conventional dead-reckoning approaches that largely suffer from drifting with the low-cost MEMS sensors, neural inertial odometry models recently emerge as a new approach to accurate inertial tracking and have been demonstrated as an effective odometry on smartphones [2, 5]. By modelling the noise and bias of low-cost inertial sensors into a supervised learning framework, such neural methods are able to yield less long-term drift in contrast to the filtering-based or double integration-based methods. Despite their success in long-distance tracking, their promise on short-term motion tracking (i.e., motion tracklets) remains unknown. Motion tracklets serve as the basic fragments of a long track followed by a moving platform. Different combination of tracklets is able to constitute or synthesize a variety of useful movements without going through every possible long-term trajectory and can be subsequently employed by different downstream applications with the minimal amount of data re-collection efforts. Unfortunately to date, such an inertial tracklet dataset together with a benchmarked performance for 6-DoF estimation (i.e., 3D position and 3D orientation) are still lacking.



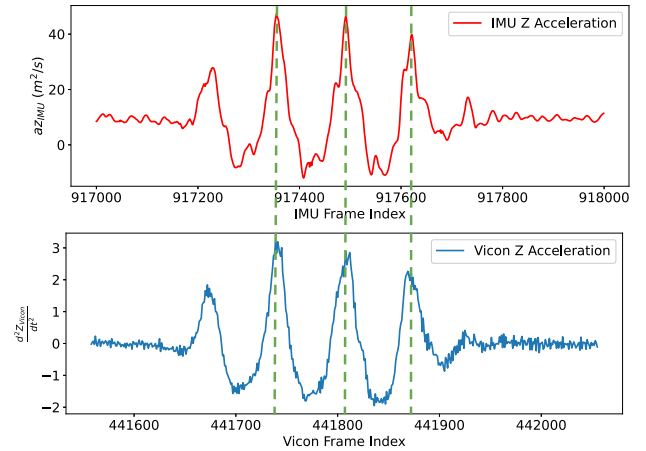
**Figure 2: Three types of tracklets recorded by our dataset: the arc-, bulge- and line-shaped tracklets. The smartphone continuously moves back and forth along its width (x-axis). Each tracklet spans around 50cm along the x-axis. Moving speed is evenly maintained so that each tracklet (left to right or right to left) takes around 5 seconds.**

In this paper, we present a new dataset for smartphone inertial motion tracking with 100 minutes of handheld motion data, designed for training neural models. The data collection setup can be seen in Fig 1. In summary, we make the following contributions:

- We present a 6DoF inertial tracking dataset with smartphone inertial data and Vicon ground truth.
- We trained a lightweight LSTM [2, 6] model to estimate the 6DoF motion tracklets. The model is benchmarked with this present dataset and is able to achieve an average Absolute Trajectory Error (ATE) of 39mm over 600mm-long tracklets. The source code and trained model are publicly released in <https://github.com/MAPS-Lab/smartphone-tracking-dataset>.
- The dataset also includes the recorded video streams of a set of ArUco markers by the smartphone camera that allows the users to expand this inertial tracking dataset to accurate visual-inertial tracking tasks.

## 2 RELATED WORK

The Oxford Inertial Odometry Dataset (OxIOD) [3] is an inertial odometry dataset that tracks 2D human/robotic motion along long distances. It aims to estimate 2D 3DoF motion in comparison to our 3D 6DoF motion. Furthermore, due to OxIOD's aim of estimating



**Figure 3: Time Synchronization using signal spikes in acceleration created by 3 consecutive jumps.**

long distances, its use cases and accuracy requirements are very different from ours, making it a dataset of a very different nature.

To the best of our knowledge, there is no dataset for accurate 6DoF inertial tracking on smartphones. Related researches collect their own data [1, 9, 10], and many of them do not have the Vicon ground truth. Some use fixed trails for the smartphone's motion [9, 10], like a toy train track or bicycle wheels. These methods can only evaluate the model by checking trajectory consistency instead of using quantitative metrics. The lack of random motion noise like in handheld motion also means the data doesn't have real-life fidelity, making it unsuitable for learning-based models. Some use advanced IMU sensors like Xsens MTx as ground truth [1], which in itself is prone to drift.

## 3 DATA COLLECTION

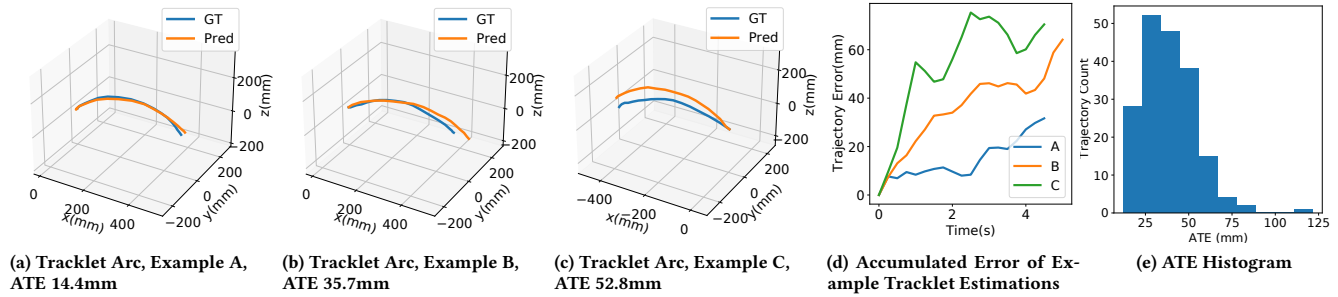
### 3.1 Collection Platform

Inertial data was collected from a Samsung Galaxy A31 smartphone. As investigated by previous literature [7], IMU sensor quality can vary between different phones, with newer or higher-end models usually providing higher-quality data. To make the dataset generalizable to more devices on the market, a mid-range Android device is specifically selected for data collection.

We collected 3 data sessions on our platform, as shown in Table 1. The third data session has an Xsens IMU, an advanced inertial sensor that produces more accurate data than smartphones. It can be used to make comparisons between low-cost smartphone sensors and industrial-grade sensors.

### 3.2 Tracklets

We carefully select three predominant types of hand motion tracklets using smartphones, including the arc, bulge and line tracks (c.f. Fig 2). As shown in Table 1, data session 1 and 2 are a mixture of all 3 types of tracklets, while session 3 is dedicated to the arc-shaped tracklet. In each data session, the tracklets are recorded in loop. The smartphone is continuously doing a back-and-forth "scanning"



**Figure 4: Benchmark results.** (a) to (d) showcases examples of our trained model estimating 6DoF tracklets using smartphone inertial data. The shown tracklets have ATE values varying from better to average to worse. (a) to (c) are the estimated 3D 6DoF tracklets compared to the ground truth, while (d) shows the accumulated drifting error of each example tracklet over motion time. (e) is the histogram of ATE on all 188 testing tracklets.

**Table 1: Meta Information of Data Sessions**

Session	Duration	Sensors	Tracklets
1	20min	smartphone, Vicon	arc, bulge, line
2	20min	smartphone, Vicon	arc, bulge, line
3	70min	smartphone, Vicon, Xsens	arc

motion along its width (left-to-right and right-to-left). Duration of each single tracklet is around 5 seconds.

Completely motionless periods are included in the data sessions as well. Including these periods in training will reduce zero-drifting produced by the model, meaning the model will not produce drifting predictions while the smartphone is in fact immobile.

### 3.3 Ground Truth

We use the Vicon tracking system as the trajectory ground truth. It provides millimeter-level position accuracy at a rate of 100Hz.

Inertial data is high-frequency and temporally sensitive, which means the Vicon ground truth and smartphone IMU need to be well-synchronized. To sync between Vicon and IMU, before and after each data session, we hold the collection rig and jump up and down 3 times. The jumps create sharp spikes in both Vicon and acceleration data. We use 3 spikes at the beginning and 3 spikes at the end of each session to temporally align IMU and Vicon signals. The finding of the spikes is done with numpy and scipy toolkits. The exact code is provided alongside the dataset.

### 3.4 Video Data

Other than inertial data and Vicon data, we also provide the video stream from the smartphone’s rear camera during dataset collection. The implication in this decision is two-fold:

- We always point the camera towards a set of ArUco markers[4, 8]. The markers could be used to generate pose estimation of the smartphone. This will enable the dataset users to generate their own ground truth, giving them the ability to

expand the dataset themselves without having access to a Vicon system.

- The video stream could potentially be used alongside inertial data to make visual-inertial odometry (VIO) systems, making it more versatile in the hands of dataset users.

## 4 BENCHMARK

We try to estimate the tracklets using neural inertial odometry: calculate the 6DoF displacement between fixed intervals and accumulate the displacement together to form a tracklet. We only train and evaluate the model on data session 3, where no mixture of different types of tracklets is present.

We use a simple LTSM model with 2 recurrent layers and 96 features in hidden state. Being only 1.2MB in size, the model is lightweight enough to be deployed on smartphone platforms. 50 consecutive IMU data frames are fed into the model, and are used to predict the relative 6DoF motion of the smartphone during these 50 frames. With IMU data frequency at 200Hz, the model predicts the smartphone’s motion at 4Hz.

For data collection Session 3, we have a total of 820,000 usable frames (68.3 minutes at 200Hz) of IMU data after time synchronization. If we predict relative 6DoF motion every 50 frames, we have 16,400 non-overlapping data samples. To further boost up the amount of trainable data, we make the starting frame of each sample increment by 10 instead of 50, overlapping each other by 40 frames. This way, IMU frame indexes of the samples are  $[[0, 49], [10, 59], [20, 69], \dots]$  instead of  $[[0, 49], [50, 99], [100, 149], \dots]$ , increasing the data samples by 5 times. In the end, we have 81,995 overlapping data samples out of the 820,000 frames of IMU data. Next, we split the data samples, making the first 65,000 out of 81,995 data samples the training set, and the last 16,995 the testing set.

The testing set is split into independent tracklets for quantitative evaluation of the estimation quality. A total of 188 separate tracklets are made from the testing set, several examples of which can be seen in Fig 4 (a)-(d). To evaluate our best model, Absolute Trajectory Error (ATE) is calculated on each tracklet, and we show a histogram of the 188 ATE results in Fig 4 (e). The mean ATE over the testing samples is 38.6mm, with a 15.3mm standard deviation.

We provide our data pre-processing, training and evaluation code as well as our best saved model alongside our dataset.

## 5 CONCLUSION

Our presented dataset provides the testing ground and metrics for smartphone inertial motion tracking that results in accurate 6DoF tracklet estimations. With neural engines and TPUs appearing on mobile platforms, the development of neural inertial tracking on smartphones can bring novel interactions and functionalities accessible to wider applications.

Meanwhile, we acknowledge some weaknesses in our dataset. We did not include high speed or sudden changes in motions, which limits the dataset to be used in a predefined range of movements. The data collection device is not diverse enough, which might caused the trained models to deteriorate on smartphones with different sensor characteristics. When the dataset is to be used in commercial rather than research purposes, the developers might need to take upon themselves to expand the data to other smartphone models and other motion tracklet sets using our ArUco marker method.

## REFERENCES

- [1] Lu Bai, Matthew G Pepper, Yong Yan, Malcolm Phillips, and Mohamed Sakel. 2020. Low cost inertial sensors for the motion tracking and orientation estimation of human upper limbs in neurological rehabilitation. *IEEE Access* 8 (2020), 54254–54268.
- [2] Changhao Chen, Xiaoxuan Lu, Andrew Markham, and Niki Trigoni. 2018. Ionet: Learning to cure the curse of drift in inertial odometry. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.
- [3] 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).
- [4] Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco José Madrid-Cuevas, and Rafael Medina-Carnicer. 2016. Generation of fiducial marker dictionaries using mixed integer linear programming. *Pattern Recognition* 51 (2016), 481–491.
- [5] Jian Gong, Xinyu Zhang, Yuanjun Huang, Ju Ren, and Yaoxue Zhang. 2021. Robust Inertial Motion Tracking through Deep Sensor Fusion across Smart Earbuds and Smartphone. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–26.
- [6] Klaus Greff, Rupesh K Srivastava, Jan Koutnik, Bas R Steunebrink, and Jürgen Schmidhuber. 2016. LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems* 28, 10 (2016), 2222–2232.
- [7] Kristian Jambrosic, Miljenko Krhen, Marko Horvat, and Tomislav Jagust. 2020. Measurement of IMU sensor quality used for head tracking in auralization systems. In *Forum Acusticum*. 2063–2070.
- [8] Francisco J Romero-Ramirez, Rafael Muñoz-Salinas, and Rafael Medina-Carnicer. 2018. Speeded up detection of squared fiducial markers. *Image and vision Computing* 76 (2018), 38–47.
- [9] Sheng Shen, Mahanth Gowda, and Romit Roy Choudhury. 2018. Closing the gaps in inertial motion tracking. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. 429–444.
- [10] Anton Umek and Anton Kos. 2014. Usability of Smartphone Inertial Sensors for Confined Area Motion Tracking.