TEXR: The University of Texas Extended Reality Headset Tracking Dataset

Robert M. Tenny
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, Texas
rmtenny@utexas.edu

Nicholas Christner
Department of Aerospace Engineering and Engineering Mechanics
The University of Texas at Austin
Austin, Texas
christner.random@utexas.edu

Raymond Jiang
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, Texas

Todd E. Humphreys
Department of Aerospace Engineering and Engineering Mechanics
The University of Texas at Austin
Austin, Texas
todd.humphreys@utexas.edu

***Abstract*—The next generation of XR devices will require seamless functionality in any environment and require globally referenced poses. This paper introduces a public benchmark dataset for extended reality (XR) headset tracking in a variety of outdoor environments. The dataset includes raw GNSS intermediate frequency samples, timestamped measurements from three inertial measurement units, and globally shuttered stereoscopic cameras, all collected on a head-mounted sensor platform. Our benchmark dataset recordings include data collected in urban, light-urban, and open field environments. The dataset includes differing levels of user dynamics corresponding to pedestrian and gaming use cases. Similar datasets exist for ground and aerial vehicle-mounted sensor platforms. However, no dataset currently exists for head-mounted XR platforms with this combination of sensors. This dataset enables the evaluation of computer vision and sensor fusion algorithms for headset tracking on real-world data.**

I. Introduction

Extended reality (XR) devices, which encompass virtual, augmented, and mixed reality devices, have become useful tools in several industries such as education, design, and entertainment [1]–[3], and the adoption of XR is expected to grow. The current generation of XR devices relies on camera-based inside-out tracking using simultaneous localization and mapping (SLAM) algorithms to determine a user headset's position and orientation (pose) relative to an ad-hoc local frame such as one arbitrarily anchored to a living room. The reliance on relative pose, as opposed to an absolute measure with respect to a well-defined global reference frame such as the International Terrestrial Reference Frame (ITRF), hinders collaboration between XR headsets in building and maintaining feature maps [4], [5]. The next generation of XR devices will be expected to function seamlessly and collaboratively in all environments, both indoor and outdoor. To bridge the gap into outdoor environments and unlock the full potential of XR, enabling natural user motion in ever-expansive virtual maps [4], [5] and high data throughput [6], precise globally referenced pose of the XR device is needed.

To develop and compare algorithms for pose estimation of XR devices, a dataset of sensor measurements from XR devices in common XR use cases is needed. This paper introduces the TExas eXtended Reality (TEXR) publicly available dataset tailored for XR headset pose estimation applications. The TEXR dataset contains raw intermediate frequency (IF) samples from two GNSS antennas, which will enable full exploration of GNSS processing techniques, GNSS observables for the dual antennas and a reference receiver, timestamped stereoscopic globally shuttered grayscale images, and inertial measurement unit (IMU) measurements from two low-cost smartphone-grade IMUs and an industrial-grade IMU. This sensor combination enables sensor fusion at the lowest level, combining vision, inertial, and GNSS data to deliver robust tracking with centimeter- and degree-level accuracy.

Several datasets exist for visual and inertial navigation. The most popular is the KITTI dataset [7], which provides a benchmark for both visual odometry and object detection for use in autonomous vehicles. Similarly, TEX-CUP [8] and UrbanNav [9] are benchmark datasets for urban positioning of ground vehicles. Both contain visual, inertial, and GNSS measurements. Although these datasets provide measurements from sensors similar to those used in TEXR, such as stereoscopic cameras and IMU measurements, the scale of the sensors is drastically different. Most notably the baseline of the stereoscopic cameras is on the scale of 1 meter, which is 10 times larger than the distance between cameras on an XR device. The dynamics also differ between ground vehicles and XR devices. Motion that may be common for an XR user, such as rapidly turning one's head left and right, is impossible for ground vehicle during normal operation. On the other hand, a ground vehicle will operate at a much higher linear (translational) velocity compared to an XR user. In

addition, XR devices and ground vehicles operate in markedly different environments. The existing ground vehicle datasets are recorded on city streets, whereas XR devices operated outdoors primarily see sidewalks, parks, and open fields. Environments such as open parks create a challenge for SLAM algorithms due to the scarcity of unique visual features. This condition is not present in vehicle datasets.

Visual navigation datasets [10], [11] for aerial vehicles and benchmark sensor fusion datasets such as the MARS-LVIG dataset [12] collected on aerial platforms contain visual, inertial, and GNSS measurements similar to TEXR. However, these datasets fall short of being fully applicable to XR positioning. Even though the scale of the sensors on aerial platforms is similar to that of an XR device, the dynamics and setting (e.g., proximity to surrounding objects) can be quite different. Most notably, an aerial vehicle has a larger range of motion in the vertical direction, while an XR device will remain in a fixed vertical range based on a user's height. Similar to ground vehicles, aerial vehicles can generally reach higher velocities compared to XR users. Another notable difference is that aerial platforms do not operate in the same types of environments as XR users; aerial platforms will often be high off the ground with navigation cameras pointed at the ground and GNSS antennas having a clear view of the sky. Further, existing datasets do not contain the case where the aerial vehicle's GNSS antennas are obstructed. In contrast, an XR device is always very close to the ground and occasionally passes under trees or bridges, obstructing the GNSS antennas.

A benchmark dataset specific to XR is necessary for the development of sensor fusion algorithms that are tailored to the unique challenges faced by future outdoor XR devices. The closest existing dataset is the HoloSet dataset [13], which contains visual and inertial data recorded by a Microsoft Hololens 2 for both the raw data and truth trajectory. The HoloSet dataset is lacking GNSS data, which limits its use for global headset tracking, and lacks trajectories in open environments with few visual features. It is paramount that pose estimation methods developed for XR devices be capable of handling challenges from open-sky conditions to deep urban canyons, from featureless open parks to feature-dense urban landscapes, and from low dynamic strolls in the park to highly dynamic XR games.

This paper and the TEXR dataset offer three primary contributions. First, they enable researchers to explore properties of XR sensors without having to create a sensor platform. Second, TEXR serves as a benchmark dataset enabling researchers to evaluate visual SLAM and sensor fusion algorithms. Third, TEXR enables other questions in XR, such as communications and scene rendering, to be explored by providing a real model of XR user dynamics. The dataset is available at https://rnl-data.ae.utexas.edu/texr/.

II. Sensor Platform

The TEXR dataset was captured using the University of Texas head-mounted sensor platform. The full data collection platform can be seen in Fig. 1.

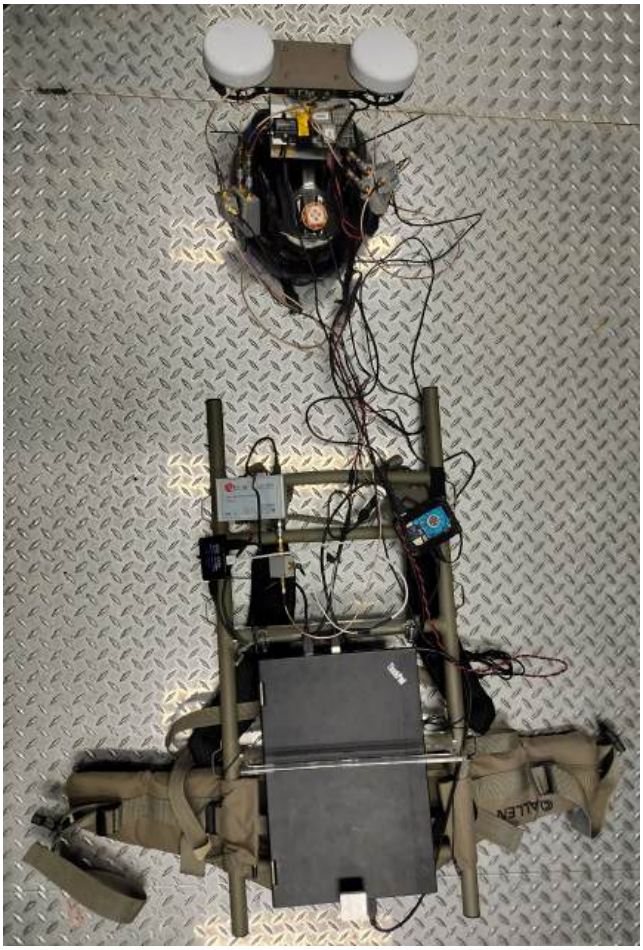

Fig. 1: The University of Texas head-mounted sensor platform for XR headset tracking research. The head-mounted sensor platform features two L1/L2/L5 GNSS antennas, two wideband GNSS RF front ends, smartphone- and industrial-grade MEMS IMUs, and stereoscopic cameras. The support hardware consisting of batteries, computers, and a GPS timing receiver is mounted to a backpack.

The sensor platform is equipped with the following sensors shown in Fig. 2:

- 2× Antcom G8Ant-3A4TNB1 1 high performance GNSS patch antennas (NGS code: ACCG8ANT 3A4TB1). Triple frequency L1/L2/L5; 40 dB low-noise amplifier
- 2× Tallysman TW1721 L1 GNSS antennas.
- 1× RadioLynx GNSS RF front end [14]–[16]. Supports two GNSS antennas; dual-frequency L1/L2; 9.6 Msps IF sampling rate on both channels; developed in-house.
- 1× RadioLion GNSS RF front end, available from Locus Lock [17], [18]. Supports two GNSS antennas; triple frequency L1/L2/L5; 10 Msps IF sampling rate on the L1 and L2 channels, 20 Msps IF sampling rate on the L5 channel; developed in-house.
- 1× u-blox EVK-M8T. Mass-market timing receiver.
- 2× Bosch BMI088 IMU. Low-cost MEMS device; smartphone-grade IMU noise characteristics; 100 Hz out-

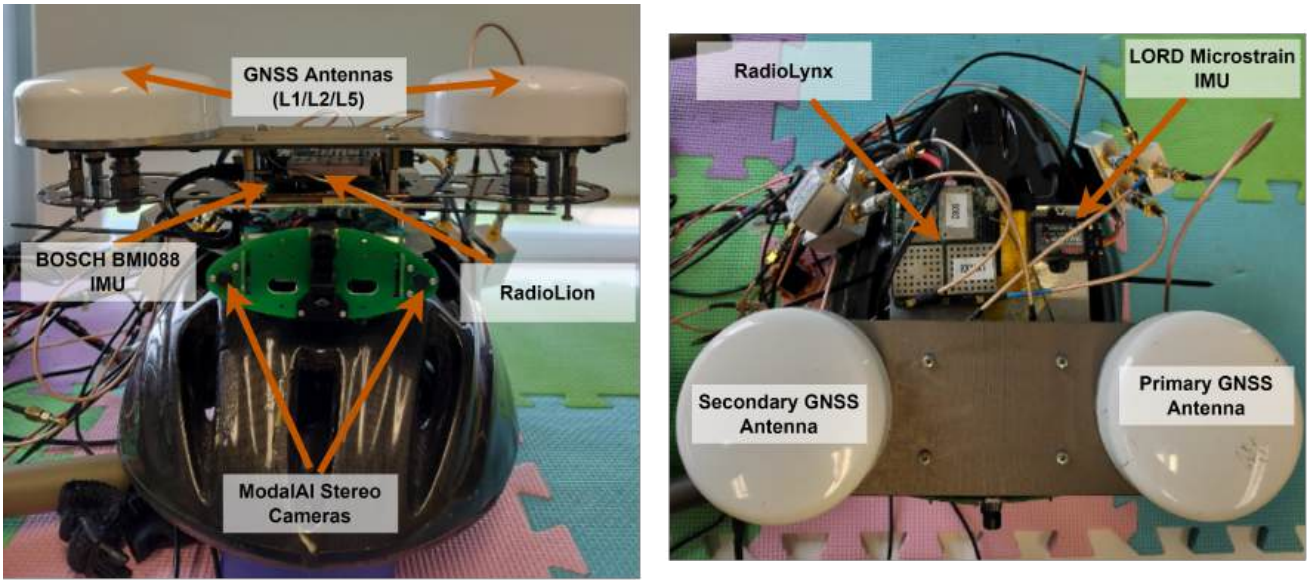

Fig. 2: The University of Texas head-mounted sensor platform contains many sensors mounted to a helmet to capture the user's motion. The exact measurements of each of the sensors' relative locations and rotations are available in the dataset.

put rate; $\pm 6\,\mathrm{g}$ specific force range; $\pm 250\,^{\circ}\,\mathrm{s}^{-1}$ angular velocity range.

- $1\times$ LORD MicroStrain 3DM-GX5-25 IMU. High-performance MEMS device; industrial-grade IMU noise characteristics; $100\,\mathrm{Hz}$ output rate.
- $2\times$ ModalAI MSU-M0015 stereo cameras. 640×480 resolution; monochromatic; OmniVision OV7251 CMOS sensor; global shutter; software-synchronized at $30\,\mathrm{fps}$;

The truth trajectory is the solution computed using the PpEngine sensor fusion engine [16]. The post-processed solution is fused carrier phase differential GNSS (CDGNSS) with inertial sensing using the Lord MicroStrain IMU and triple frequency GNSS measurements from the GRID software-defined receiver, whose application name is PpRx [14], [16], [18], [19].

The head-mounted sensor platform backpack houses a Lenovo Yoga, and Odroid XU4 computers. The Lenovo Yoga logs data from the RadioLion, Bosch BMI088, and LORD MicroStrain IMU as shown in the Fig. 5. The ModalAI VOXL, housed on the headset, logs data from the stereo cameras. The Odroid XU4 logs data from the RadioLynx and provides timestamps for the images recorded by the ModalAI VOXL as shown in Fig. 5.

Most data logging processes are developed in-house for precise synchronization between sensor data. Details on sensor synchronization are provided in Sec. IV.

To enable CDGNSS-based positioning, the TEXR dataset also includes GNSS data logged from a nearby reference antenna with a clear view of the sky. The reference antenna is a geodetic-grade Trimble Zephyr II (NGS code: TRM57971.00). For consistency with the headset, reference data are logged with an identical RadioLynx RF front end. The sensor platform is always within $2\,\mathrm{km}$ of the reference antenna, representing short-baseline CDGNSS conditions.

The Cartesian coordinates (in meters) of the reference antenna's L1 phase center in the ITRF Earth-Centered Earth-Fixed (ECEF) reference frame at the time of the data capture are

$$\boldsymbol{r}_{\mathrm{ECEF}} = [-742080.0705, -5462030.6811, 3198339.7301]^{\mathsf{T}}$$

III. Data Collection

The TEXR dataset captures a mix of open-sky and light-urban environments to represent a mix of outdoor XR use cases. XR devices will be used in two main regimes, *pedestrian* and *gaming*. Trajectories in the *pedestrian* case model a engaging in video calls, navigation, or messaging. In the *pedestrian* cases a user walking with slow head motion deviating from the direction of travel models a user occasionally looking around. The trajectories often explore new areas in a mix of open-sky and light-urban environments. Several data recordings exist for both the open-sky and light-urban portions of the University of Texas campus. An example of a *pedestrian* trajectory is shown in Fig. 3. The dataset contains a mix of closed-loop trajectories such as Fig. 3 and non-closed trajectories so that algorithms that implement loop closure can be tested.

Data recordings for the *gaming* case have different user dynamics compared to the *pedestrian*. In the *gaming* trajectories a user's translation and rotation is faster compared to the *pedestrian* trajectories. The user is moving at a quick walking or jogging pace with rapid head motions and the user's head is often not facing the direction of travel. The *gaming* trajectories are in open-sky conditions while the user stays in a local region such as an open field and the user often revisits previous locations. An example of the *gaming* case is shown in Fig. 4.

Fig. 3: Example of *pedestrian* headset trajectories from the dataset. These trajectory were recorded on the University of Texas at Austin campus open-sky conditions. The user is facing the direction of motion during the majority of the trajectory. The top trajectory represents an open-loop case where the user returns to a previous location. The lower image is an example of an open-loop trajectory.

A KML file is provided along with each recording for easy visualization.

The dataset captures a variety of open-sky and light-urban environments in order to enable the testing of visual SLAM algorithms in realistic scenarios. There is a complementary trade-off between GNSS and visual sensing in open-sky and light-urban environments. Open-sky environments such as open fields lack visual features while also having many GNSS satellites in view. The light-urban trajectories have fewer GNSS satellites in view while also having a larger number of visual features due to the surrounding buildings.

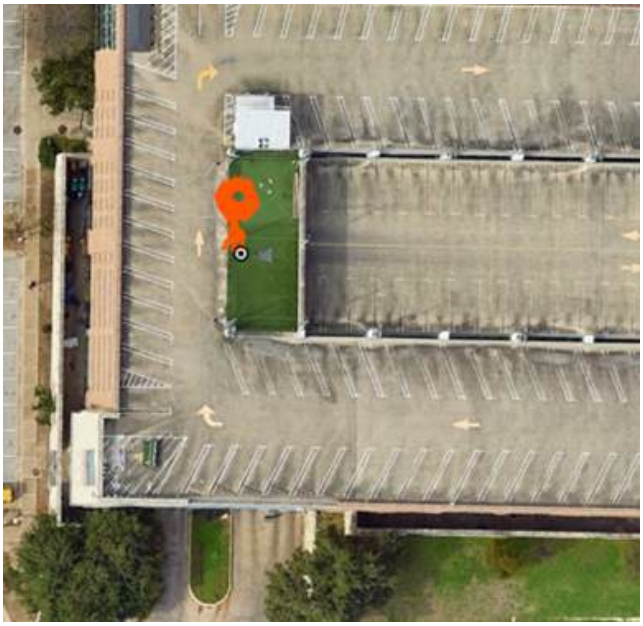

Fig. 4: Example of a *gaming* headset trajectory under open-sky conditions recorded on the roof of the Speedway parking garage. In this trajectory the user remains in the same local area. The user moves at a higher rate in both translation and rotation compared to the *pedestrian* trajectories. The user is not always looking in the direction of travel.

The mix of trajectories provides a challenge for headset tracking algorithms to achieve robust performance in all outdoor environments.

A. Data Formats

This section describes the data formats available for the data collected by each sensor. The descriptions are organized by sensor.

1) RadioLynx: The RadioLynx RF front end generates single-bit-quantized samples from two antennas on the headset and a single antenna at the reference station, capturing $10\,\mathrm{MHz}$ bandwidth at both L1 and L2 GPS bands. The dataset provides tracked L1 pseudorange and carrierphase observables generated by the GRID software-defined receiver [16]. The RadioLynx is used to provide timestamps for the stereo images. At the time of writing, GRID tracks GPS, Galileo, BeiDou, and SBAS signals. The observables are provided in the RINEX format.

2) RadioLion: The RadioLion RF front end generates dual-bit-quantized samples from two antennas at the headset capturing $10\,\mathrm{MHz}$ bandwidth at both L1 and L2 bands and $20\,\mathrm{MHz}$ bandwidth around the GPS L5 frequency. The raw IF data from the two antennas are made available in a binary format documented along with the dataset [17], including the required IF parameters. Raw IF data enable development of new signal tracking strategies for precise positioning, and allow high-sensitivity receivers to track weak signals that may not have been tracked by the receivers in the recording platform. Raw

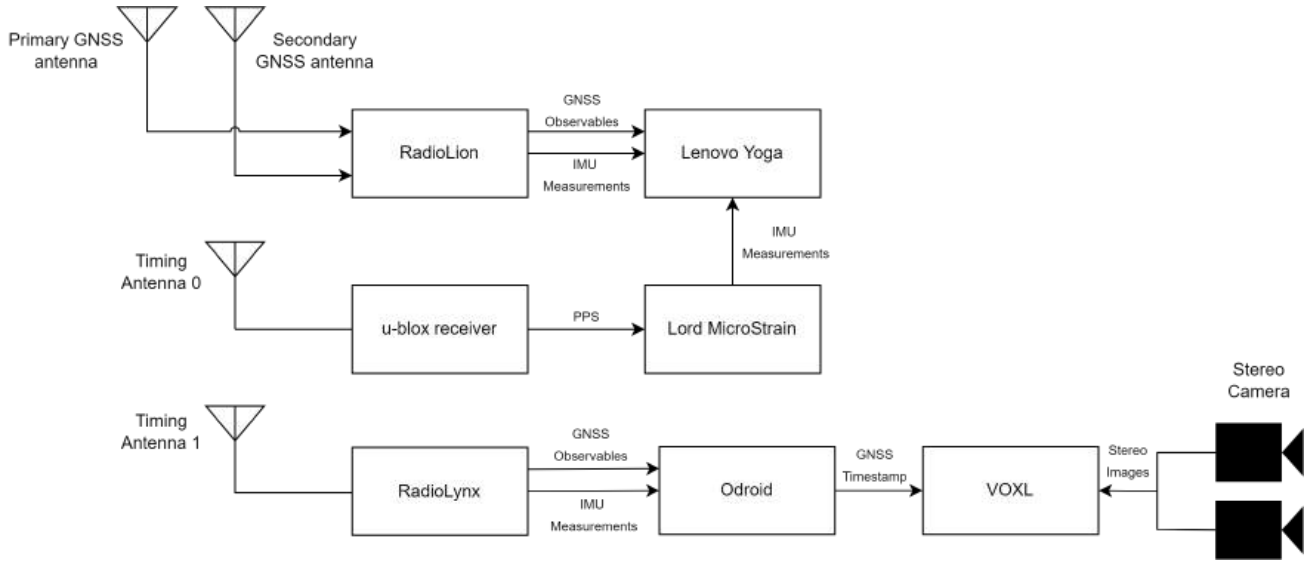

Fig. 5: Block diagram of the sensors and computers on the University of Texas head-mounted sensor platform. The platform consists of four main sensors and three computers so that the data can be recorded quickly with timestamps that can be directly traced back to GPS time.

IF samples from the reference antenna can be used for data bit wipeoff [14] if desired. The dataset also provides tracked pseudorange and carrierphase observables generated by the GRID software-defined receiver [16].

3) Stereo Cameras: The timestamped stereo images from the ModalAI cameras are saved in the HDF5 format. The HDF5 file contains images saved as PNG files and timestamps given as the GPS nanosecond of the week as described in Sec. IV. The images used for camera calibration are provided in the same HDF5 format along with the description of the calibration chart used to calibrate the cameras.

4) Bosch IMU: The Bosch BMI088 IMU data are recorded at $100\,\mathrm{Hz}$ and timestamped with the GPS time of each measurement. The data are available in a CSV format. The TEXR dataset provides the extrinsic IMU parameters, lever arm and rotation in the head-mounted sensor platform's body frame. A description of the synchronization and calibration is found in Sec. IV. The IMUs are built-in to the RadioLynx and RadioLion boards, and have been set up such that the IMU data timestamps can be traced back to the GNSS RF sampling clock. This enables highly accurate correspondence between the IMU timestamps and GPS time.

5) Lord Microstrain IMU: Timestamped specific force and angular rate measurements from the high-performance LORD MicroStrain MEMS IMU are made available in CSV format. The LORD IMU accepts a PPS (pulse per second) signal generated by the u-blox receiver to synchronize to GPS time. LORD IMU measurements are internally compensated for temperature variation.

6) Ground Truth Trajectory: The ground truth trajectory is provided in a CSV format and as a KML file. The ground truth trajectory is found using the GNSS observables from the RadioLion and the LORD Microstrain IMU. The truth trajectory contains a pose solution found by performing a fused GNSS-INS position and orientation estimation using the PpEngine sensor fusion engine [16].

IV. Sensor Calibration and Synchronization

Accurate sensor calibration and synchronization is imperative for any localization dataset. The performance of the headset tracking is highly dependent on the precision of the calibration of the sensors. This section describes the extrinsic and intrinsic calibration performed for each of the sensors. The relative positions and orientations of the sensors are shown in Fig. 6. The dataset contains all the files used to calibrate the sensors so that the calibration can be repeated.

A. IMU calibration

The lever arm between the the phase center of the primary GNSS antenna and the IMUs is provided with the orientation of each of the IMUs and is shown in Fig. 6. The dataset provides a long calibration trajectory under open-sky conditions with reference observables for refinement of the extrinsic IMU parameters. A $24\,\mathrm{h}$ long stationary recording for the Bosch and Lord IMUs is also provided in the dataset to enable the calibration of the noise and bias stability. The accelerometer and gyroscope biases and scale factors can also be estimated from these data, however these values should be treated as an initialization for an online calibration as they will change due to the turn-on bias of the IMUs.

B. Camera Calibration

The lever arm between the the phase center of the primary GNSS antenna and the cameras is shown in Fig. 6 and is provided with the orientation of each of the cameras. A series of images of a calibration chart and the calibration chart parameters are provided in the dataset so that calibration can

be performed. The intrinsic parameters found using the Kalibr calibration toolbox [20] are provided in the dataset.

C. Camera Synchronization

The ModialAI stereo cameras are triggered according to the VOXL's internal clock at a rate of $30\,\mathrm{fps}$. Each of the globally shuttered image sensors are triggered simultaneously. At the start of each exposure the current GPS timestamp is recorded as nanoseconds of the week. The timestamp is pulled from the Odroid XU4 running the GRID software defined receiver as shown in Fig. 5. The mean latency between the recorded timestamp of the exposure time and the true exposure time was found to be less than $1.5\,\mathrm{ms}$. The latency was found by recording an LED powered by the PPS output of the u-blox EVK-M8T timing receiver. The LED illuminated precisely at the start of the GPS second. The time stamp of the image with the first illumination of the LED differenced from the known time to find the error in the timestamp. It was found that the error in the timestamp was less than the exposure time of the camera.

V. Summary and Future Extensions

The TEXR dataset provides a benchmark for testing and evaluating headset tracking algorithms in real outdoor environments. The authors hope this comprehensive dataset of GNSS observables and raw IF data, IMU measurements, and stereo camera images benefits futures researchers. The TEXR dataset enables the evaluation of computer vision and sensor fusion algorithms on real world data and become a benchmarking service for XR similar to the KITTI dataset [7] for computer vision on vehicles. The authors plan to expand the dataset in the future to include trajectories in deep-urban environments, and trajectories with minimal GNSS availability. The TEXR dataset will also be expanded to include a 5G/6G cellular FR3 receiver to enable research into improving wireless connectivity for XR devices.

Acknowledgment

Research was sponsored by affiliates of the 6G@UT center within the Wireless Networking and Communications Group at The University of Texas at Austin.

References

[1] S. Zhang and W. Li, "Applying extended reality (XR) technology in commerce, management, and business applications: A survey," in *2024 4th International Conference on Computer, Control and Robotics (ICCCR)*. IEEE, 2024, pp. 108–113.

[2] A. Becker and C. M. D. S. Freitas, "Evaluation of XR applications: A tertiary review," *ACM Computing Surveys*, vol. 56, no. 5, pp. 1–35, 2023.

[3] S. Dargan, S. Bansal, M. Kumar, A. Mittal, and K. Kumar, "Augmented reality: A comprehensive review," *Archives of Computational Methods in Engineering*, vol. 30, no. 2, pp. 1057–1080, 2023.

[4] T. E. Humphreys, R. X. T. Kor, and P. A. Iannucci, "Open-world virtual reality headset tracking," in *Proceedings of the ION GNSS+ Meeting*, Online, 2020.

[5] R. M. Tenny, L. C. Sun, A. Duru, and T. E. Humphreys, "Robust absolute headset tracking for extended reality," in *Proceedings of the IEEE/ION PLANS Meeting*, Monterey, CA, 2023.

[6] A. Duru, M. Mozaffari, M. Afshang, T. Zhang, T. Khan, T. E. Humphreys, and J. G. Andrews, "Pose-aware 3D beamwidth adaptation for mobile extended reality," in *IEEE International Conference on Communications*. IEEE, 2024.

[7] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *International Journal of Robotics Research (IJRR)*, 2013.

[8] L. Narula, D. M. LaChapelle, M. J. Murrian, J. M. Wooten, T. E. Humphreys, J.-B. Lacambre, E. de Toldi, and G. Morvant, "TEX-CUP: The University of Texas Challenge for Urban Positioning," in *Proceedings of the IEEE/ION PLANS Meeting*, 2020.

[9] L.-T. Hsu, F. Huang, H.-F. Ng, G. Zhang, Y. Zhong, X. Bai, and W. Wen, "Hong Kong UrbanNav: An open-source multisensory dataset for benchmarking urban navigation algorithms," *NAVIGATION*, vol. 70, no. 4, 2023.

[10] W. Xu, Y. Yao, J. Cao, Z. Wei, C. Liu, J. Wang, and M. Peng, "UAV-VisLoc: A large-scale dataset for UAV visual localization," *arXiv preprint arXiv:2405.11936*, 2024.

[11] A. George, N. Koivumäki, T. Hakala, J. Suomalainen, and E. Honkavaara, "Visual-inertial odometry using high flying altitude drone datasets," *Drones*, vol. 7, no. 1, p. 36, 2023.

[12] H. Li, Y. Zou, N. Chen, J. Lin, X. Liu, W. Xu, C. Zheng, R. Li, D. He, F. Kong *et al.*, "MARS-LVIG dataset: A multi-sensor aerial robots SLAM dataset for LiDAR-visual-inertial-GNSS fusion," *The International Journal of Robotics Research*, p. 02783649241227968, 2024.

[13] Y. Chandio, N. Bashir, and F. M. Anwar, "Holoset-a dataset for visual-inertial pose estimation in extended reality: Dataset," in *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, 2022, pp. 1014–1019.

[14] T. E. Humphreys, M. J. Murrian, and L. Narula, "Deep-urban unaided precise global navigation satellite system vehicle positioning," *IEEE Intelligent Transportation Systems Magazine*, vol. 12, no. 3, pp. 109–122, 2020.

[15] J. E. Yoder, P. A. Iannucci, L. Narula, and T. E. Humphreys, "Multi-antenna vision-and-inertial-aided CDGNSS for micro aerial vehicle pose estimation," in *Proceedings of the ION GNSS+ Meeting*, Online, 2020, pp. 2281–2298.

[16] J. E. Yoder and T. E. Humphreys, "Low-cost inertial aiding for deep-urban tightly-coupled multi-antenna precise GNSS," *NAVIGATION*, vol. 70, no. 1, 2023.

[17] Z. Clements, P. A. Iannucci, T. E. Humphreys, and T. Pany, "Optimized bit-packing for bit-wise software-defined GNSS radio," in *Proceedings of the ION GNSS+ Meeting*, St. Louis, MO, 2021, pp. 3749–3771.

[18] H. A. Nichols, M. J. Murrian, and T. E. Humphreys, "Software-defined GNSS is ready for launch," in *Proceedings of the ION GNSS+ Meeting*, Denver, CO, 2022.

[19] T. Pany, D. Akos, J. Arribas, M. Z. H. Bhuiyan, P. Closas, F. Dovis, I. Fernandez-Hernandez, C. Fernández-Prades, S. Gunawardena, T. Humphreys, Z. M. Kassas, J. A. L. Salcedo, M. Nicola, M. L. Psiaki, A. Rügamer, Y.-J. Song, and J.-H. Won, "GNSS software defined radio: History, current developments, and standardization efforts," *NAVIGATION*, vol. 71, no. 1, 2024.

[20] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 1280–1286.

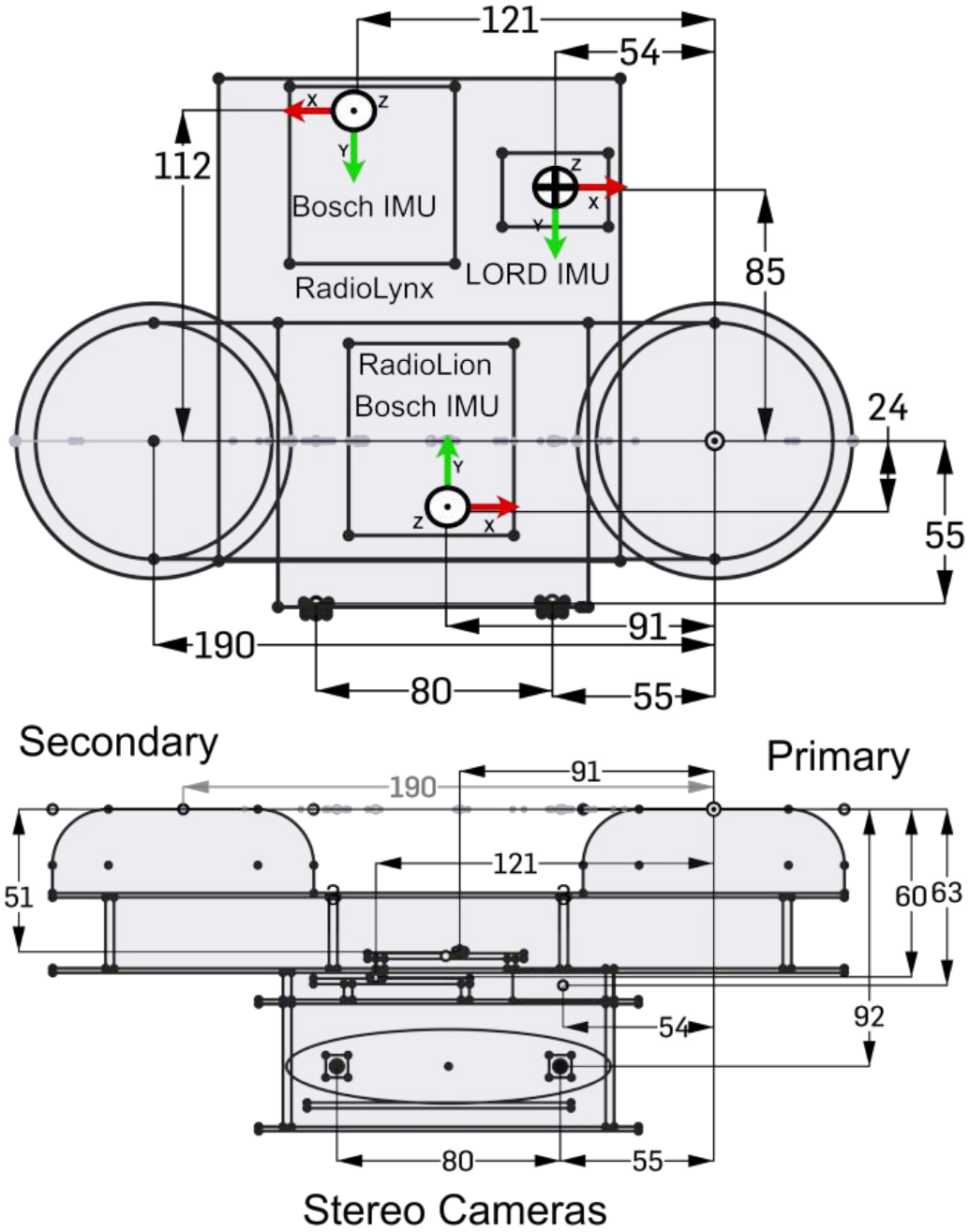

Fig. 6: Top: top-down view of the sensing platform. Bottom: front view. The drawings indicate the relative positions and orientations of all the sensors on the head-mounted platform. All the measurements are in millimeters. The relative rotation of each of the three IMUs is shown in the top panel by denoting the xyz coordinate system of each IMU. The symbol $\odot$ denotes vectors coming out of the page while the symbol $\otimes$ denotes a vector going into the page. All coordinate systems are right-handed and orthogonal.