

Indoor positioning systems: Smart fusion of a variety of sensor readings

Maximilian Arnold

Nokia Bell Labs, Germany

Stuttgart, Germany

maximilian.arnold@nokia-bell-labs.com

Frank Schaich

Nokia Bell Labs, Germany

Stuttgart, Germany

frank.schaich@nokia-bell-labs.com

ABSTRACT

Robust and versatile localization techniques are key to the success of the next industrial revolution. Yet, it is uncertain which combination of sensors will be the most robust and valuable. Thus, we present a versatile and reproducible measurement system incorporating a manifold number of state-of-the-art sensors to compare and fuse the raw input data. It is shown that some techniques show very good results on the same scenario and data-set, but fall apart on translating to a slightly different scenario. In general we show that the vanilla approach to fuse the raw data achieves reasonable results in the generalization domain, demonstrating that radio frequency (RF) localization techniques in combination with an inertial measurement unit (IMU) could result in a very robust and promising candidate for solving this challenging task.

1 INTRODUCTION

Robust and versatile indoor positioning systems (IPs) are key to the success of the next industrial revolution [1, 2]. It can be seen as an key enabler for a wide range of applications such as indoor navigation, smart factories, or it could even provide a basic security functionality in distributed Internet of Things (IoT) sensor networks [3–5]. In general IPs will incorporate a manifold of different sensors, to enhance the different systems and creating a practical system. Although IPs are currently holding back fully autonomous factories due to their reliability regarding their accuracy, it is uncertain which combination of sensors will be the most efficient in terms of cost and usability.

Different robust IPS techniques have emerged ranging from trilateration [6] over angle-of-arrival [7] to recent deep learning (DL) approaches [8]. All of them demonstrate promising results in large parts of a given area, but suffer in specific parts of the environment. Therefore approaches to fine-tune Neural Networks (NNs) [9] or advanced fusion techniques [10, 11] were proposed.

Yet, to the best of our knowledge we are the first to present a data-set containing a manifold of off-the-shelf RF, magnetic and IMU localization techniques being compared in a fair,

and neutral manner allowing us to analyze and rate the different techniques. Thus, this work emphasises on *how* to automatically generate a large data-set in different scenarios and compare various approaches for fusion. We demonstrate the feasibility of synchronizing the data over Network Time Protocol (NTP) and Message Queuing Telemetry Transport (MQTT) protocols and verify the sanity of our measurement data using classical localization techniques. Moving towards robust positioning systems a vanilla state-of-the-art DL approach on each of the modalities demonstrate the applicability of these systems. Further we consider a raw-data fusion method using DL for enhancing the overall localization performance in all scenarios.

Due to their underlying structure, DL approaches tend to focus heavily on a specific scenario, losing generality and the ability to be transferred to other domains. Thus, we measure two separate data-sets within one day, with a slightly different setup and try to predict from one to the other. We demonstrate that this is in general possible but with a certain amount of loss.

To sum up this paper main contributions are:

- presenting a method to create automatically a large amount of labeled data in a plug and play fashion, extending to a manifold of sensors
- a data-set containing synchronized sensor readings for various techniques ultra-wideband (UWB)-range, channel state information (CSI), received signal strength indicator (RSSI), magnetic information, Odometry data
- IMU is shown to be a very valuable element for raw-data fusion, as its independent of the environment
- CSI and IMU data fused together show the best generalisability and enabling efficient IPS.

2 POSITIONING SYSTEMS

Tackling the challenging task of indoor localization a manifold of techniques have been proposed, built and tested, yet a clear vision, as e.g. available in outdoor scenarios like the global positioning system (GPS), is missing. These proposed techniques can be classified into three classes: (1) RF-based localization techniques, e.g. CSI or UWB; (2) Inertial tracking,

e.g. exploiting inertial sensors and (3) camera/infrared/laser based techniques, e.g. Light Detection and Ranging (LiDaR).

Resulting from the fact that IoT devices demand battery lives over multiple months/years, these techniques have to be lightweight or computational heavy algorithms have to be shifted out, e.g. to an edge cloud, allowing for more complex and diverse techniques, ranging from trilateration over fingerprinting techniques to advanced NN structures.

However, all of these techniques suffer from different deficiencies; RF based techniques can experience outages, while inertial tracking drifts over time and camera/LiDaR systems are expensive and suffer under low lighting conditions and blockage. Thus, in general a robust Simultaneous Localization and Mapping (SLAM) algorithm using multiple sensors to create precise maps is missing. To tackle this challenge multiple fusion approaches were envisioned ranging from Kalman filter over particle filter to conditional random fields, all allowing to fuse the data in a raw or processed format. As localization algorithms typically are hand-crafted towards their application, the common method is to fuse the data not in the raw but in a pre-processed fashion. We consider here the typical, yet broadly used, classical techniques and introduce a straightforward and simple fusion method at the raw data level.

Trilateration

Trilateration was one of the first localization techniques allowing to pinpoint a user in a mobile network with a coarse accuracy. This technique is commonly used in GPS devices and has been heavily improved over time.

Fig. 1 depicts the basic concept of trilateration. Hereby the three distances d_0, d_1, d_2 are estimated through measuring time of flights between the target and the different anchors. The estimated distances can be represented as circles around the anchors, which intersect in the best case in only a single point, the target location x, y . For a more simplified solution the anchor 0 is used as reference point, e.g. (x_0, y_0) is set to 0 by shifting the remaining coordinates by the point, resulting in

$$\begin{aligned} d_0^2 &= x_{\text{target}}^2 + y_{\text{target}}^2 \\ d_1^2 &= (x_{\text{target}} - x_1 - x_0)^2 + (y_{\text{target}} - y_1 - y_0)^2 \\ d_2^2 &= (x_{\text{target}} - x_2 - x_0)^2 + (y_{\text{target}} - y_2 - y_0)^2 \end{aligned}$$

As the distance measurements are noisy the circles usually are not intersecting in a single point and thus a different approach needs to be applied, by first intersecting two circles; forming a line, which can be intersected with the remaining circle. This, will execute trilateration well, even if the circles do not intersect perfectly. *Fingerprinting:*

Fingerprinting is a method to localize a target in a well defined-environment, where a map is created in an initial

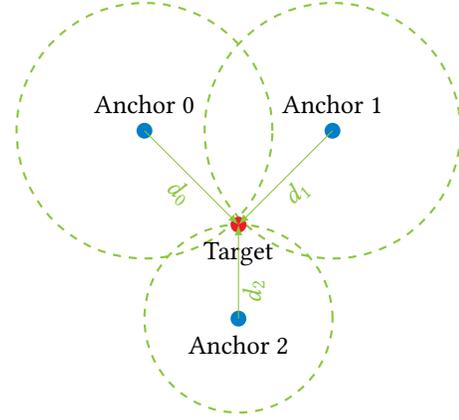


Figure 1: Trilateration example

training phase to be used in the deployment phase to locate the target. Thus, a cumbersome training phase creating an N dimensional vector per position is key for the success of this technique. In the deployment phase the measured N dimensional vector is correlated to the data set. Then either the closest trained vector or an interpolation between the closest trained vectors is used to estimate the position. Due to the huge amount of overhead, this technique is slow, very susceptible to changes of the environment and not considered in most cases.

Deep Learning positioning:

DL techniques are getting more and more applied in many different applications ranging from image classification to speech detection. Its main feature is to map an input space to an output space over a regression or classification task, where classical models are either too complex, too susceptible for different impairments or too computational heavy. Thus in this case a non-linear transformation

$$\mathbf{m} \xrightarrow{f_{\Theta}} (\hat{x}, \hat{y}),$$

has to be learned, which maps the input measurement vector \mathbf{m} to a discrete position, depending on the hyperparameters Θ . Thus allowing for a system learning the environment and the underlying structure of the measurements.

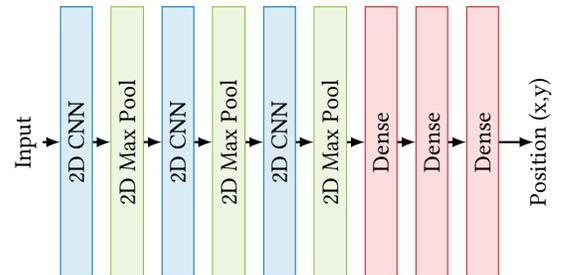


Figure 2: NN architecture optimized and searched via AutoML tools.

Fig. 2 depicts the NN architecture searched via the auto machine-learning (AutoML) tool from Microsoft [12]. The hyperparameters were also optimized via this tool, allowing for a defined comparison for each technique¹. This NN will be used throughout the remaining paper with optimized kernel/pool/stride/dense sizes based on the input measurements using the tool [12].

3 MEASUREMENT SYSTEM

The success for DL in vision and audio based techniques was achieved only by leveraging the large datasets being already available for classical algorithms. For creating a highly used data-set for IPS, we introduce a versatile measurement setup, allowing us to add in a plug-and-play fashion any kind of sensor.

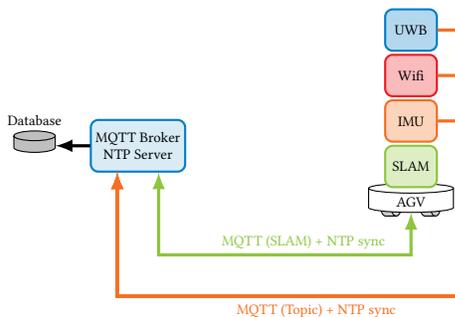


Figure 3: Versatile measurement system.

Fig. 3 depicts the flexible and versatile measurement platform for automatically creating a large amount of labelled data. It is based on a vacuum cleaner robot including a two-dimensional LiDaR, ultrasonic distance sensors, wheel-tick sensors, multiple infrared (IR) sensors, and an IMU. These modalities are fused together—via a Kalman filter-based SLAM algorithm—to generate pseudo-groundtruth for our subsequent experiments. The SLAM accuracy is better than 1 cm, as per the evaluation conducted by Hoffmann et al. [6]. This level of accuracy satisfies the requirements of most localization applications.

In general we are creating the system by using: (1) an robot for the pseudo-groundtruth, (2) a platform mounted on top of the robot to host additional sensors, (3) a MQTT broker with a database for data acquisition, and (4) an NTP server for time synchronisation between various sensors. In terms of hardware, the platform was fabricated inhouse from 3D printed parts and a LEGO building plate. On the software side, a MQTT broker and a NTP server form the core of the data acquisition system. Specifically, each sensor synchronises its time-clock with the local NTP server, which allows for an accuracy below 1 ms [13]. The synchronised measurements

¹The hyper-parameters and layer sizes will be distributed with the dataset

are then streamed to the MQTT broker. By virtue of MQTT and the back-end database, our sensor system is modular and extensible.

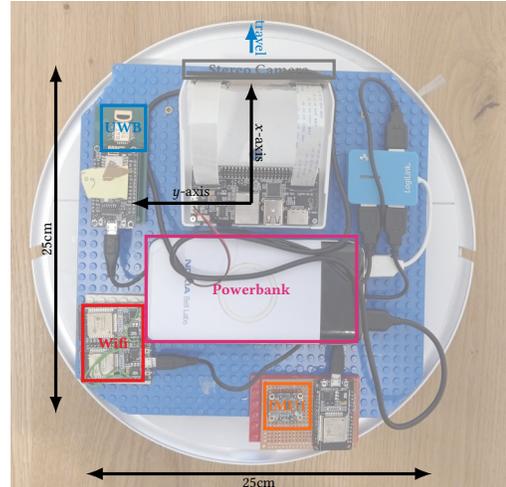


Figure 4: Build up of measurement platform consisting of: A roborock vacuum cleaner; a building platform; a modality of sensors and a power-bank.

We leverage a Roborock S50 from Xiaomi as base robot platform shown in the build-up figure 4. The robotic platform streams its pseudo-ground-truth positions to the MQTT broker with a 5 Hz update rate. A power-bank of 10Ah is used to power the sensors. We utilise IMU modules from Adafruit [14] equipped with Bosch BNO055 inertial chipsets [15]. Each IMU module is connected to an ESP32 2.4 GHz WiFi chip-set over an I2C interface. The UWB system consists of three anchors and one tag using the DW1000 chip-set. All of them are using the band 5 with 500 MHz bandwidth. The WiFi CSI and RSSI is achieved by using two ESP32 listening to the pilot WiFi signals of the surrounding anchors. The anchors are all ESP32 creating traffic by constantly pinging the Fritzbox router.

3.1 Pseudo groundtruth

Fig. 5 depicts the measurement area within a flat’s floor-plan located in Stuttgart, Germany. The robot trajectories are highlighted in blue. The floor-plan has been automatically created by the robotic SLAM algorithm as described above. The measurement campaign was conducted over a single day creating one data-set in the morning and another data-set in the evening. Coordinate translation of the SLAM pseudo ground-truth is needed in order to compensate for all sensor positions, as the robot can spin around at a certain location moving the sensors on a circular trajectory. In order to correct for these effects, we devised the following. Firstly, the coordinate system of the SLAM calculates the position based

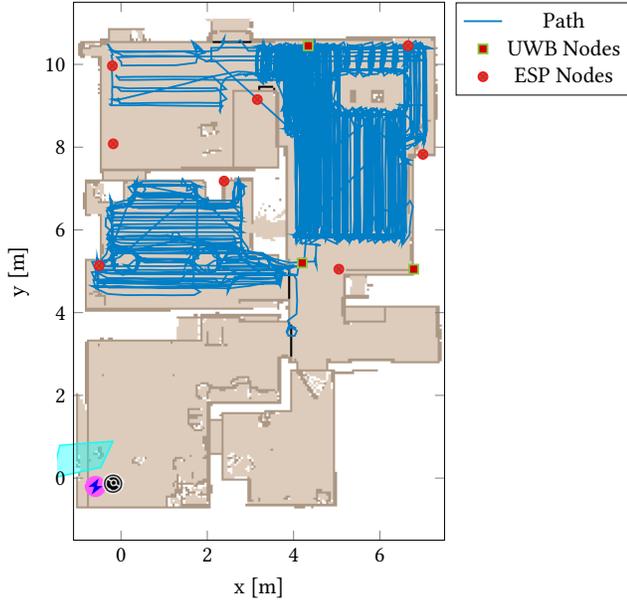


Figure 5: Measurement paths and anchor placements in the flat.

on an anchor point located at the front of the robot. The anchor point is translated to the geometrical center of the robot. The SLAM ground-truth coordinates $(x_{\text{SLAM}}, y_{\text{SLAM}})$ and orientation ϕ_{SLAM} are referenced to the robot coordinate system. Denoted by $(\hat{x}_{\text{SLAM}}, \hat{y}_{\text{SLAM}})$ and $\hat{\phi}_{\text{SLAM}}$ we get a translated reference coordinate and orientation system, respectively. Then the offset angle $\phi_{\text{Sensor}, \ell}$ and coordinates $(x_{\text{Sensor}, \ell}, y_{\text{Sensor}, \ell})$ of the sensor were measured. Finally, the measured values are used to translate the position of the sensor to the global coordination system, according to

$$\begin{aligned} r_{\text{Sensor}, \ell} &= \sqrt{x_{\text{Sensor}, \ell}^2 + y_{\text{Sensor}, \ell}^2} \\ \hat{x}_{\text{Sensor}, \ell} &= \hat{x}_{\text{SLAM}} + \text{Re} \left\{ r_{\text{Sensor}, \ell} e^{j(\hat{\phi}_{\text{SLAM}} + \phi_{\text{Sensor}, \ell})} \right\} \\ \hat{y}_{\text{Sensor}, \ell} &= \hat{y}_{\text{SLAM}} + \text{Im} \left\{ r_{\text{Sensor}, \ell} e^{j(\hat{\phi}_{\text{SLAM}} + \phi_{\text{Sensor}, \ell})} \right\} \end{aligned} \quad (1)$$

Tab. 1 depicts the two available measurement data-sets used for analyzing sensor fusion and positioning algorithms. We have repeated the same measurement twice to understand if the algorithms generalize or not. Thus, a system trained on Dataset 1 should be able to infer the position on the Dataset 2. The difference between the two data-sets is not exactly the same measurement through slightly changing the building plate while adding additional sensors. This will give a strong indication on the robustness and versatility of the system.

Parameter	Dataset 1	Dataset 2
Time Run	160.17 min	141.0 min
Path Covered	1975.28 m	1733 m
Data	Morning	Afternoon
Nb. CSI Anchor	13 (ESP)	13 (ESP)
Nb. UWB Anchor	3 (DW1000)	3 (DW1000)
Nb. IMU	1 (Bosch BNO05)	1 (Bosch BNO05)
Nb. Cameras	0	1 (Pi Stereo Hat)
Labels	2D LiDaR	2D LiDaR
Update Rate CSI	7.50 Hz	8.5 Hz
Update Rate UWB	9 Hz	7 Hz
Update Rate IMU	76.93 Hz	76.93 Hz
Nb points CSI	34417	46584
Nb. points UWB	52152	28239
Nb. points IMU	747853	656640
Data ESP	CSI + RSSI	CSI + RSSI
Data UWB	Range + Power	Range + Power
Data IMU	9-DoF +Odo	9-DoF +Odo

Table 1: Two measured datasets used for localization fusion.

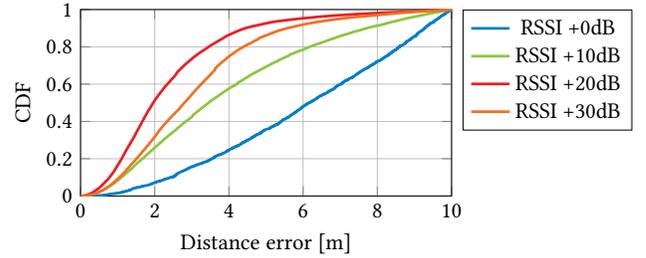


Figure 6: Importance of RSSI calibration for trilateration based on distance.

Emphasizing that this is raw data directly streamed from the sensor, we highlight the necessity of calibration and fusing the sensors correctly, like in an actual deployment (comp. Fig. 6). Hereby the RSSI per node was used to estimate the distance to the target and the three closest RSSI sources were used to do trilateration. It turns out that the RSSI in this setup performs the best if every RSSI measurement is enhanced by 20dB allowing for a more robust trilateration.

4 RESULTS

A practical localization system in general requires to achieve: (1) at least in 99% an distance error to be smaller than 1 m; (2) being able to handle small changes in the environment, e.g. translating from one data-set to another. First, the general performance of the single algorithms is evaluated; then we investigate sensor fusion and finally we analyze the generalizability over time.

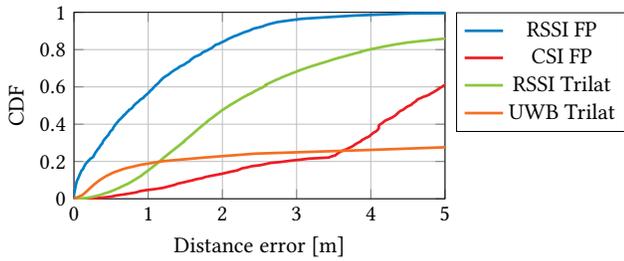


Figure 7: Performance of classical methods on the Dataset 1.

Fig. 7 the cumulative distribution function (CDF) of the distance error for several variants of a fingerprinting-based (FP) method and two trilateration methods. It becomes apparent that due to the limited number of UWB anchors the trilateration succeeds only in 20 % of the cases, indicating there is an area with a higher accuracy while the remaining area is only poorly covered by three nodes. The RSSI trilateration using the virtue of more nodes achieves a better accuracy, yet still only in a similar range as commercial GPS devices. The CSI fingerprinting methods fails, as the phase of the anchors is not fixed and thus an arbitrary phase shift between nodes is created. An advanced classical methods might be able to track this, but this would exceed the scope of this paper. The RSSI FP method gives a good indication what can be achieved with a higher number of nodes in the UWB system. This indicates a high likelihood for fusion of methods to join the advantages of these methods and remove unambiguous features.

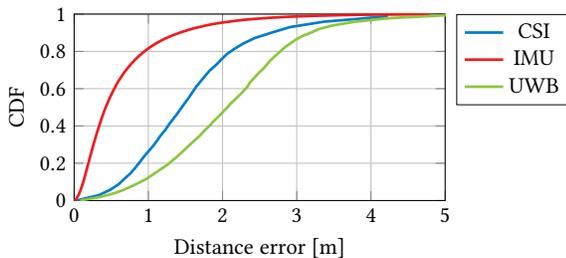


Figure 8: Performance of vanilla NN on the different raw inputs from Dataset 1.

Fig. 8 depicts the results of the vanilla NN optimised using Neural Network Intelligence (NNI), where the CSI, IMU (3D magnetic, 3D acceleration, 3D gyro) and UWB (range) data is feed into the NN and optimized using a 90/10 split between training and testing. It becomes apparent that the UWB performance is better due to the ability to filter out points with only one connect node allowing to use the middle as the estimate position, thus limiting the error. Moreover the CSI method seems to be achieving better performance than its classical pendant the fingerprinting method, due to more information being available and finding a common

function for the features. Due to the magnetic fingerprinting, the performance of the vanilla IMU NN seems to be very solid. Note that tracking in this case is not possible, as the data is shuffled. The tracking ability will be used in the last section, by slicing a certain amount of time together. In general the techniques achieve similar performances as the classical methods but removing the outliers.

4.1 Fusion

To remove drift and to enhance performance the raw data of the various techniques is stacked together and the effective performance of a vanilla NN is estimated.

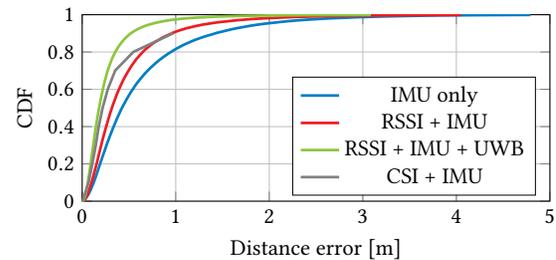


Figure 9: Performance of the vanilla NN on the stacked raw data.

Fig. 9 includes the IMU curve as reference (blue). Each addition of further sensor data improves the system performance as expected. In general the RSSI in combination with the IMU data shows potential in removing the outliers in the IMU system. The CSI in combination with the IMU shows also reasonable results in average below than 10 cm and in 95% better than 1 m accuracy. Yet, the best system comes from the highest dimensional input, where the RSSI is combined with the IMU and UWB system; allowing even in 99% of the cases to be closer than 1 m. It becomes apparent that joining sensors even in this vanilla case allows for a better performance. Note: This is on the same data-set thus it does not give an indication on the generalizability, which will be investigated next.

4.2 Time stability (generalizability)

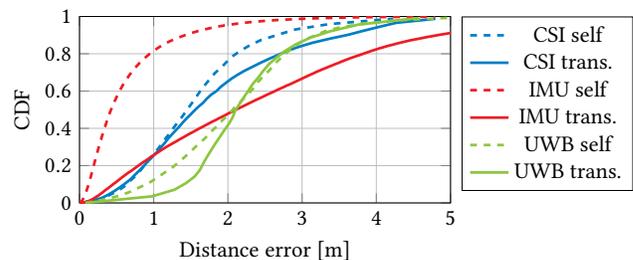


Figure 10: Performance of the vanilla NN from the morning to the evening data-set.

Fig. 10 demonstrates the performance of the NN being trained on the Dataset 1 and evaluated on the Dataset 2. The solid curve gives the self performance meaning on the same 90/10 split dataset and the dashed curve gives the generalizability by evaluating without taking any data from Dataset 2 the performance. We show that the IMU system performs very well being on exactly the same configuration, but small changes in the position or the rotation removes any advantage of the system. So a better generalization method needs to be applied to make IMU data viable in itself. It can be seen that the UWB system also did some form of over-fitting on the actual data as it has a better lower part of the CDF, but yet it achieves similar performance on the overall data-set, showing that it generalized well. The best generalization is achieved by the CSI approach where no over-fitting happened, but only a small degradation of the curve occurred. It shows that the measurement is reproducible and viable to do such kind of experiments. The results indicate that we can use a combination of sensor readings to achieve a reasonable performance level.

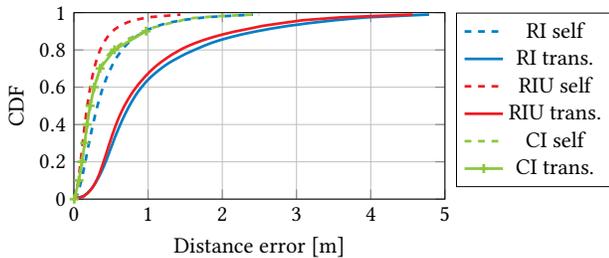


Figure 11: Performance of the fused methods trying to predict from one data-set to another.

Fig. 11 depicts the performance of the fused techniques. All RSSI based techniques although having performed well on the first data-set, would require additional labels to stabilize performance. Thus, these methods require continuous updates to adjust to small changes in the scenario. In the case of the CSI-based method in combination with the IMU raw data it turns out that the curves exactly match, which was already hinted in the previous results, where the drift of the CSI was not as impacted as the other techniques. Thus the combination of the IMU and CSI data seem to be a solid combination for a reliable and robust positioning technique, achieving on average around 10 cm of accuracy and in the 95% around 1m accuracy.

5 CONCLUSION

In this paper we have presented a versatile and robust measurement system for creating large amount of labeled data for the investigation of localization techniques. We demonstrate that this dataset is viable for different fusion and mapping techniques. Different classical localization methods are compared with state-of-the-art NN approaches demonstrating

that the NN techniques can handle outliers better than pre-defined algorithms. A vanilla approach to fuse the raw-data outputting the position is proposed. The CSI data combined with the IMU data have shown high performance, indicating the combination of those techniques to be a generalized high quality localization technique.

REFERENCES

- [1] R. Ayyalasomayajula, A. Arun, C. Wu, S. Sharma, A. R. Sethi, D. Vasisht, and D. Bharadia, "Deep learning based wireless localization for indoor navigation," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–14.
- [2] M. Brossard, A. Barrau, and S. Bonnabel, "Ai-imu dead-reckoning," *IEEE Transactions on Intelligent Vehicles*, 2020.
- [3] C. Chen, Y. Chen, Y. Han, H. Q. Lai, F. Zhang, and K. J. R. Liu, "Achieving Centimeter-Accuracy Indoor Localization on WiFi Platforms: A Multi-Antenna Approach," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 122–134, Feb 2017.
- [4] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "ConFi: Convolutional Neural Networks Based Indoor Wi-Fi Localization Using Channel State Information," *IEEE Access*, vol. 5, pp. 18 066–18 074, 2017.
- [5] N. Lee and D. Han, "Magnetic indoor positioning system using deep neural network," in *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Sept 2017, pp. 1–8.
- [6] D. Hoffmann, F. Fellhauer, M. Arnold, and S. ten Brink, "Indoor Measurements with Commercial 60 GHz WLAN Components," in *WSA 2020; 24th International ITG Workshop on Smart Antennas*, 2020, pp. 1–5.
- [7] Z. H. Wu, Y. Han, Y. Chen, and K. J. R. Liu, "A Time-Reversal Paradigm for Indoor Positioning System," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1331–1339, April 2015.
- [8] M. Arnold, J. Hoydis, and S. t. Brink, "novel massive mimo channel sounding data applied to deep learning-based indoor positioning," in *SCC 2019; 12th International ITG Conference on Systems, Communications and Coding*.
- [9] M. Arnold, S. Dörner, S. Cammerer, J. Hoydis, and S. ten Brink, "Towards Practical FDD Massive MIMO: CSI Extrapolation Driven by Deep Learning and Actual Channel Measurements," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 1972–1976.
- [10] S. Gunaga, N. C. Iyer, and A. Kulkarni, "Sensor Fusion Based State Estimation for Localization of Autonomous Vehicle," in *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, ser. AutomotiveUI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 89–91.
- [11] D. Weber, C. Gühmann, and T. Seel, "Neural networks versus conventional filters for inertial-sensor-based attitude estimation," 2020.
- [12] Microsoft, "NNI (Neural Network Intelligence) is a lightweight but powerful toolkit to help users automate Feature Engineering, Neural Architecture Search, Hyperparameter Tuning and Model Compression." 2021. [Online]. Available: <https://github.com/microsoft/nni>
- [13] D. L. Mills, "Internet time synchronization: the network time protocol," *IEEE Transactions on Communications*, vol. 39, no. 10, pp. 1482–1493, 1991.
- [14] Adafruit, "ADAFRUIT 9-DoF." [Online]. Available: <https://www.adafruit.com/product/2472>
- [15] Bosch, "BNO055 Intelligent 9-axis absolute orientation sensor," 2020. [Online]. Available: https://cdn-shop.adafruit.com/datasheets/BST_BNO055_DS000_12.pdf