CONVERGE Challenge: Multimodal Learning for 6G Wireless Communications

ICASSP 2026 SP Grand Challenge

Guidelines for Task 2 – User Equipment Localization and Position Prediction

v1.0 (22/12/2025)

Jichao Chen
Communication Systems Department, EURECOM, France
jichao.chen@eurecom.fr

Filipe B. Teixeira
INESC TEC, Faculdade de Engenharia, Universidade do Porto, Portugal
filipe.b.teixeira@inesctec.pt

Francisco M. Ribeiro
INESC TEC, Faculdade de Engenharia, Universidade do Porto, Portugal
francisco.m.ribeiro@inesctec.pt

Ahmed Alkhateeb
Arizona State University, Tempe, USA
alkhateeb@asu.edu

Luis M. Pessoa
INESC TEC, Faculdade de Engenharia, Universidade do Porto, Portugal
luis.m.pessoa@inesctec.pt

Dirk Slock
Communication Systems Department, EURECOM, France
dirk.slock@eurecom.fr

December 22, 2025

Abstract

The CONVERGE Challenge: Multimodal Learning for 6G Wireless Communications aims to leverage synchronized visual and wireless data to support communication tasks such as blockage prediction, UE localization, channel prediction, and beam prediction. This report describes the blockage prediction task, the experimental setup used for data collection, the dataset structure and annotations, and the evaluation protocol and participant guidelines.

Contents

List of Figures

List of Tables

Chapter 1

Task Overview

1.1 Description

The UE Localization and Position Prediction task focuses on estimating the current position of the UE and predicting its future position in a dynamic mmWave/FR2 communication environment. Accurate UE position information is a key enabler for beam management, channel prediction, and proactive resource allocation in high-frequency wireless systems.

1.2 Objective

The main objective of this task is to predict the UE position at t ms for each video frame.

Chapter 2

Experimental Setup

2.1 Chamber Setup

To collect the data for this task, we have setup the CONVERGE Chamber in an environment similar to the one illustrated in Fig. 2.1.

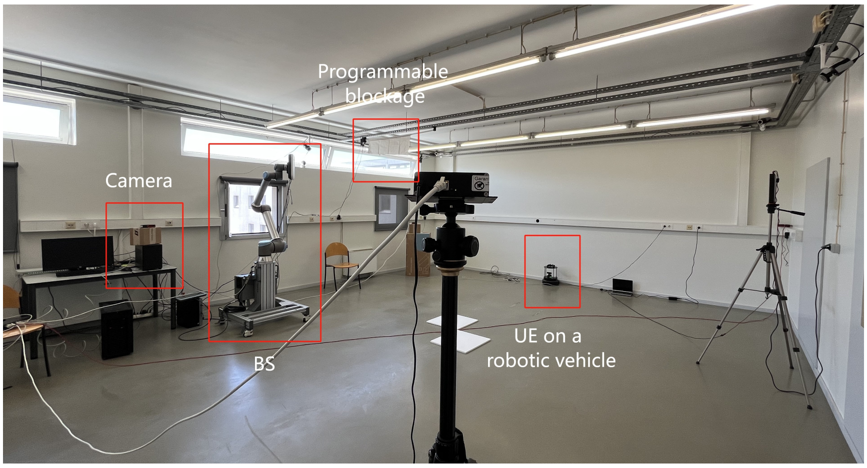

Figure 2.1: CONVERGE chamber with mobile BS, UE, and camera in Porto, Portugal.

2.2 Hardware Components

- **Mobile Base Station (BS)**: Mounted on a Universal Robotics UR-10e robotic arm. Implemented with a LiteOn FR2 Radio Unit as the mmWave front-end, integrated with the OpenAirInterface (OAI) software stack to implement the 5G NR gNB. The robotic arm is programmed to do left/right movements to create Line-of-Sight (LoS) / Non-Line-of-Sight (NLoS) conditions.

- **Mobile User Equipment (UE)**: RM530F UE from Quectel, mounted on a Turtlebot 4 robot, fitted with mmWave phased arrays oriented in multiple directions.
- **Nerian 3D Camera**: Precision Time Protocol (PTP) synchronized Nerian Ruby 3D RGB-D camera positioned at the BS to capture the surrounding environment and UE interactions.
- **Qualisys Motion Tracking System**: Provides sub-mm tracking with markers mounted on top of the UE and at the front plane of the RU.

2.3 Experiments

The data was collected using the CONVERGE software stack and is divided into five experiments.

2.3.1 Exp 6. UE on Turtlebot - random movement within FoV

The UE in mounted on the Turtlebot and moves in a random pattern within the camera Field of View (FoV). Fig. 2.2 shows the mobile gNB setup with the camera mounted on top of the RU and Fig. 2.3 shows the mobile User Equipment (UE) with the markers on top.

Figure 2.2: Detail of th mobile gNB.

Figure 2.3: Mobile User Equipment (UE) mounted on a Turtlebot 4.

2.3.2 Exp 7. UE on Turtlebot - random movement within FoV

This experiment is another run of Exp. 6.

2.3.3 Exp 8.UE on Turtlebot - random movement starting and finishing out of the FOV

In this experiment, the UE starts its movement outside of the Field of View (FoV) of the camera mounted on the RU. Then, it enters the FoV and moves randomly, before goig out of the FoV.

Chapter 3

Dataset

3.1 Modalities

For each experiment, we collected radio and video data.

3.1.1 Radio data

Radio data is collected via (1) the E2 interface and (2) SRS measurements.

E2 interface Provides PTP-synced data every 50 ms, including parameters such as: ue_id, cellid, in_sync, rnti, dlBytes, dlMcs, dlBler, ulBytes, ulMcs, ulBler, ri, pmi, phr, pcmax, rsrq, sinr, rsrp, rssi, cqi, pucchSnr, puschSnr, dlQm, ulQm.

SRS data Captured every 10 ms for each antenna (0 and 1). The estimated channel is obtained via Least Squares (LS) estimation by dividing received signal by the pilot reference, followed by interpolation to filter noise and estimate subcarriers without an associated pilot. Given 200 MHz bandwidth, 128 PRBs are used. Multiplying PRBs by 12 yields 1584 channel estimates per frame, per antenna, every 10 ms.

3.1.2 Video data

Video is captured at 7 fps (one frame every 142 ms), from the Radio Unit (RU) towards the UE. The camera is mounted on top of the RU.

3.2 Ground Truth Annotations

Ground truth is provided by manual annotation CSV that contains per-sample translation of Quectel w.r.t. liteon: timestamp,x,y,z (mm), expressed in the liteon frame `annotations/` folder.

3.3 Dataset Layout

```
dataset/
  calibration/nerian_gnb_1_calib.yaml
  task1/exp1..exp5/
    annotations/*.csv
    radio/E2-*.csv
    radio/SRS-*.json
    video/frames.csv
    video/color/
    video/disparity/
  index.csv
```

The `nerian_gnb_1_calib.yaml` provides a camera calibrationt hat allows 3D reconstruction from disparity frames.

3.4 Index File

`dataset/index.csv` has one row per scenario with dataset-relative paths.

Columns

- `task`: `task1, task2 or task3`
- `scenario_id`: `exp1..exp8`
- `scenario_name`: short label
- `video_frames_csv`: path to frame index CSV
- `video_color_dir`: path to color frames directory
- `video_disparity_dir`: path to disparity frames directory
- `radio_e2`: path to E2 CSV
- `radio_srs`: path to SRS JSON
- `annotation`: path to annotation CSV
- `notes`: short scenario description

3.5 Timestamps

- Video timestamps are in seconds (float).
- Radio timestamps are in seconds (float) and preserve absolute time. The radio files were normalized to seconds (suffix `_sec`).
- Annotation timestamps are in seconds (float).

3.6 Task 2 Annotation Format

Annotation CSV contains per-sample translation of Quectel w.r.t. liteon: timestamp,x,y,z (mm), expressed in the liteon frame. The file name varies per experiment; use `index.csv` to locate it.

3.7 Notes on `frames.csv`

`video/frames.csv` uses paths relative to the `video/` directory:

```
color/img_XXXX.png
disparity/img_XXXX.png
```