User identification by matching radio “vision” and computer vision through means of machine learning

Vinicius M. de Pinho[1] and Dalia Popescu[1]

[1]Nokia Bell-Labs, 91620 Nozay, France

***Abstract*—In the context of 5G networks, the possibility to fusion "radio vision" with "computer vision" is a must-have asset, enabler of building the extensive navigation maps through the so-called MirrorWorld. This demo will showcase an essential building block for merging the two environments: matching a user equipment’s identity in video stream and in radio measurements.**

We demonstrate the integration of a computer vision system with a radio access network and showcase the identification of the true radio transmitter between two equipment existing in a video feed.

I. Introduction

Within the broad spectrum of application and scenarios of the fifth generation (5G), the Digital Twin [1] stands out a novel application with tough requirements driven by the harsh demands of ultra-reliable and low latency (URLLC) users, yet with a enlarged variety of possibilities to explore in setups with high flexibility and increased sensing capabilities like the industrial private networks. In this context, achieving a MirrorWorld [2] becomes foreseeable through combining the reliability of radio systems and the advanced perception of the space of industrial sensors, all enabled through the intelligence of machine learning.

Projects like ARENA2036 [3] or Nokia-Omron [4] partnership are already trying to bring Industry 4.0 [5] setups to reality in a speedy fashion. One of the direct targets of this merging is user positioning and tracking. While user positioning through exploiting radio capabilities has been done with quite resource demanding algorithms and methods [6], [7], video recognition provides a broad perspective with advanced capabilities of tracking and computing. Some recent works have looked into bringing intelligence from computer vision systems to the radio networks [8], [9] yet the question of matching the identity of an equipment in technologies that sense the space in a fundamentally different manner stays opened.

In this sense, **our demonstration is as following**. We showcase a test-bed that comprises a computer vision (CV) system and three radio devices. One of the devices acts as an access point (AP), one as a connected user equipment (UE) and one as a rogue user equipment, not transmitting and not connected to the AP. By exploiting information from video feed and the radio signature produced by the devices and by means of machine learning (ML), the system associates the identity of a user in the radio domain to a transmitting device identified in the video domain.

Annex to ISBN 978-3-903176-28-7© 2020 IFIP

II. Framework Description

The computer vision system generates a video feed towards the computer to which the radio devices, universal software radio peripherals (USRPs) Ettus B210, are connected (see Figure 1). The test-bed comprises 3 USRPs that act as AP, user device and rogue device, respectively. The connected USRP sends a pilot-based frame with a BPSK modulation using an orthogonal frequency-division multiplexing (OFDM) transmission at 1 GHz.

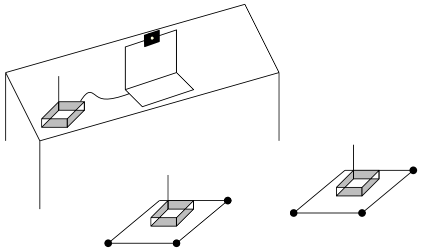
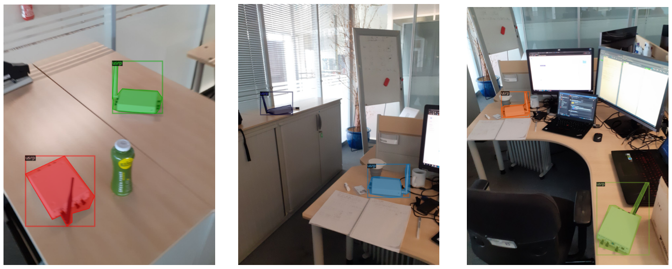


Figure 1: Setup

The visual detection is done by fine-tuning a Mask region-based convolutional neural network (R-CNN) model available in the Detectron2 framework [10] to recognize a USRP and output a bounding box (BBOX). Figure 2 shows USRPs and their BBOXs used to fine-tune the model.



(a) Example 1.

(b) Example 2.

(c) Example 3.

Figure 2: Examples of annotated images with bounding boxes used for training the model.

The computer vision system produces a list of BBOXs identifying UEs in the video feed. The AP receives from the connected UE radio frames and computes the channel impulse response (CIR). We design a ML system that gets as input a list of "B" BBOXs produced by the CV system and channel estimations for a targeted UE attached to the AP. The ML’s objective is to output which one of the spotted UEs in the video feed is the one transmitting.

The first step is data acquisition, illustrated in Figure 3.

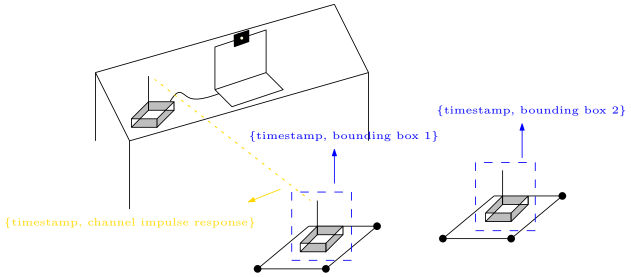

{timestamp, bounding box 2}
{timestamp, bounding box 1}
{timestamp, channel impulse response}

Figure 3: Data acquisition.

As the goal is to identify which of two UEs is transmitting, we collect video feed containing two UEs, therefore two BBOXs, and radio frames from a singular transmitting UE. For both UEs, the positions are varied through space. As data from the vision and radio domains are acquired concurrently yet with a considerable different periodicity, we use timestamps as unique identifiers to match the video and radio measurements and generate an unified structure with CIRs and BBOXs.

Next procedure is the feature extraction, as depicted in Figure 4.

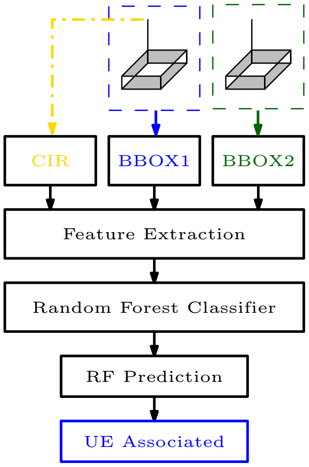


Figure 4: Illustration of the classification pipeline.

The features extracted from the radio domain are all CIR related: the CIR magnitude, phase, the value and index of the CIR magnitude peak in the radio frame, the mean value and standard deviation of the CIR magnitude vector. The features from the visual domain are the BBOXs acquired with the fine-tuned model from Detectron2. To train the classification model, the complete set of features are fed to the classifier.

We use a Random Forest classifier (RFC), which is an ensemble learning algorithm that uses N classification trees to make a decision [11]. Each tree process a subset of the available set of features and states a prediction, the N predictions are used to establish the RFC prediction.

Figure 5 shows the input to the RFC, along with the label. Each input instance has its own label as $X \in \mathcal{X} = \{0, 1, 2\}$, where $X = 0$ represents for no UE is transmitting, $X = 1$ means that the UE from BBOX 1 is the transmitting and $X = 2$ for UE from BBOX 2 is transmitting.

Features				Label
CIR	CIR related	BBOX 1	BBOX 2	X

Figure 5: Classifier's input instance.

The model training is carried out combining an exhaustive search over RFC parameters values, for the best number of trees and the best maximum depth of the trees. The training data represents 80% of total amount of the data and the validation 20%. The training uses 10-fold cross-validation procedure and two different metrics are used for evaluation in each iteration, the logarithmic loss and the F_1 score. The best model is selected and used for validation, where we compute the confusion matrix, precision, recall, F_1 score and classification accuracy.

Figure 6 shows intuitively the setup of our demo.

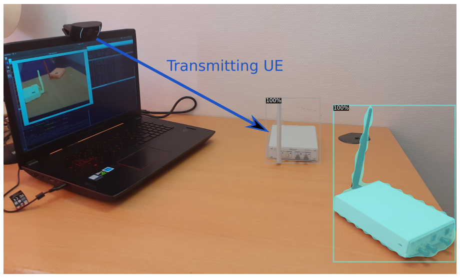


Figure 6: Demo idea

References

[1] D. Hartmann and H. van der Auweraer, "Digital twins," 2020.

[2] Kevin Kelly, "AR Will Spark the Next Big Tech Platform—Call It Mirrorworld." [Online]. Available: https://www.wired.com/story/mirrorworld-ar-next-big-tech-platform/

[3] "ARENA 2036," 2020. [Online]. Available: https://www.arena2036.de/en/

[4] "DOCOMO to Commence 5G Trials at Manufacturing Sites in Partnership with OMRON and Nokia." [Online]. Available: https://www.nttdocomo.co.jp/english/info/media_center/pr/2019/0910_00.html

[5] "Industrie 4.0." [Online]. Available: fr.wikipedia.org/wiki/Industrie_4.0

[6] Y. Lee and et al., "Implementation of multiple signal classification and triangulation for localization of signal using universal software radio peripheral," *2019 IEEE 4th ICCCS 2019*, 2019.

[7] A. Sobehy and et al., "NDR: Noise and Dimensionality Reduction of CSI for Indoor Positioning Using Deep Learning," in *2019 IEEE GLOBECOM*, Dec 2019, pp. 1–6.

[8] M. Alrabeiah and et al., "Viwi: A deep learning dataset framework for vision-aided wireless communications," in *submitted to IEEE Vehicular Technology Conference*, 2019.

[9] M. Alrabeiah and et al, "Millimeter Wave Base Stations with Cameras: Vision Aided Beam and Blockage Prediction," in *submitted to IEEE Vehicular Technology Conference 2019.*

[10] Y. Wu and et al, "Detectron2," github.com/facebookresearch/detectron2.

[11] D. Denisko and et al, "Classification and interaction in random forests," *Proceedings of the National Academy of Sciences*, vol. 115, no. 8, pp. 1690–1692, feb 2018.