WiP Abstract: Deep Intelligent Network for Device-free People Tracking

Yang Zhao
GE Research
Niskayuna, New York, USA

Ming-Ching Chang
University at Albany, SUNY
Albany, New York, USA

Peter Tu
GE Research
Niskayuna, New York, USA

ABSTRACT

Recent radio frequency (RF) sensing techniques use a network of RF sensors to detect and locate people that do not carry any devices and can operate in non line-of-sight environments. Model-based device-free RF sensing systems use statistical models to quantify human presence and motion based on the received RF signal measurements. However, such methods often require the fine tuning of multiple model-dependent parameters in order to achieve sub meter accuracy. In this work, we propose to use deep neural networks together with visual tracking systems to effectively generate training data so as to learn a general model. Our method can automatically produce human motion and occupancy images from RF sensor network measurements without the need for manual RF model parameter tuning.

KEYWORDS

Deep Neural Networks, Detection, Tracking, RF Sensor Network.

ACM Reference Format:
Yang Zhao, Ming-Ching Chang, and Peter Tu. 2019. WiP Abstract: Deep Intelligent Network for Device-free People Tracking. In *10th ACM/IEEE International Conference on Cyber-Physical Systems (with CPS-IoT Week 2019) (ICCPS '19), April 16–18, 2019, Montreal, QC, Canada.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3302509.3313312

1 INTRODUCTION

As billions of sensors and embedded devices are deployed and interconnected ubiquitously, tremendous amounts of data are available for artificial intelligence (AI) algorithms to find data patterns and predict trends. Deep Neural Network (DNN) based learning methods have been successful in many fields including object detection, recognition and human activity recognition [1] *etc.* In this work, we propose to apply DNN to an important Internet of Things (IoT) application — device-free people tracking with radio frequency (RF) sensor networks [3, 4].

A network of low-cost, RF sensors can detect and track human motion and presence in real-time even in non line-of-sight (NLOS) environments [3]. Since this technology does not require any person-borne devices, it is called sensorless or device-free tracking, and it has many applications including security, smart facilities, building energy management, emergency first response, search and rescue, *etc.* However, many state-of-the-art device-free people tracking systems require domain experts to configure and calibrate the systems before accurate detection and localization performance can be achieved. For example, the work of [4] requires an expert to manually tune six parameters for the radio tomographic imaging (RTI) system to achieve sub meter accuracy in a multipath-rich environment. To remove the human expert in the loop, we propose to use: (1) supervised deep learning methods together with (2) automatic labeled data generation using visual detection and tracking systems [2], to effectively learn a generic RF sensing model for device-free people detection and tracking.

ICCPS '19, April 16–18, 2019, Montreal, QC, Canada

ACM ISBN 978-1-4503-6285-6/19/04.
https://doi.org/10.1145/3302509.3313312

2 METHODS

Figure 1 shows the overall architecture of the proposed people tracking system. Computer vision (CV) systems can detect and track people in well-lit conditions in line-of-sight (LOS) environment. Recent visual object detection algorithms (YOLOv3, SSD, Faster-RCNN, or Mask R-CNN) can accurately detect and localize people when they are visible. On the other hand, RF sensing can augment visual sensing and compensate occlusion. We propose to combine the CV systems [2] and RF sensor networks [4] to build the next-generation indoor site-wide people detection and tracking system for occupancy estimation in a real-world settings. The combination of the two sensing modalities can increase coverage, bootstrap each other allowing for improved detection accuracy and reduced errors. Furthermore, CV systems can be used to create ground truth data for automatic RF sensing parameter tuning.

For NLOS environments, we propose to design new experiments with CV systems and adopt existing NLOS experiments [3, 4] to obtain location ground truth data for data-driven parameter learning. With the learning of both LOS and NLOS environments, a generic model can be built to operate in diverse settings and environments.

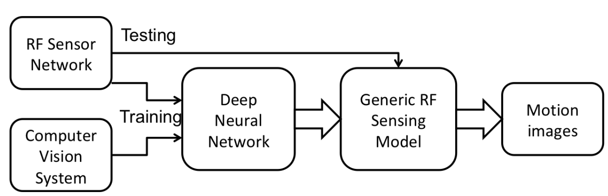

Figure 1: System Architecture.

2.1 Device-free RF tracking system

Unlike RFID localization systems, device-free RF sensing systems directly use the attenuation, reflection and scattering effects of the human body on radio signals to detect and localize people. Thus they do not require person-borne devices. In this paper, we focus on kernel distance-based radio tomographic imaging (KRTI) [4], a model-based device-free tracking system. The KRTI system uses low-cost radio transceivers as sensors to reconstruct an occupancy image (examples shown in Figure 2). Essentially, the KRTI algorithm estimates the occupancy image $\hat{\mathbf{x}}$ using the kernel distance $\mathbf{d}$ calculated from RF measurements:

$$\hat{\mathbf{x}} = (W^T C_{\mathbf{n}}^{-1} W + C_{\mathbf{x}}^{-1})^{-1} W^T C_{\mathbf{n}}^{-1} \mathbf{d}, \tag{1}$$

where $C_{\mathbf{x}}$ is the covariance matrix of the occupancy $\mathbf{x}$ (human presence), $C_{\mathbf{n}}$ is the covariance matrix of the measurement noise, and W is the forward model [3]. The occupancy covariance can be modeled as the following [4]:

$$\left[\frac{1}{\sigma_n^2} C_{\mathbf{x}}\right]_{i,j} = \frac{\sigma^2}{\delta} \exp\left(-\frac{\|\mathbf{r}_j - \mathbf{r}_i\|}{\delta}\right), \tag{2}$$

where $\sigma^2 = \sigma_x^2/\sigma_n^2$ is the ratio of variance of occupancy σ_x^2 to the variance of noise σ_n^2 which is used as a regularization parameter, δ is a correlation distance parameter, $\mathbf{r}_i$ and $\mathbf{r}_j$ are the center coordinates of the i-th and j-th pixels. We will show in § 2.4 the effects of the covariance parameters σ_x^2 and δ *w.r.t.* the occupancy images. Note that once the occupancy images are reconstructed, person locations and counting can be estimated from the resulting heatmap images.

2.2 Computer vision tracking system

Visual detection from low-cost cameras can provide accurate people localization in the LOS environments. We propose to leverage visual person detection and tracking systems in two ways: (1) A computer vision system can be used to gather ground-truth data for automatic RF parameter learning. (2) The combination of visual detection and RF detection systems can jointly increase coverage and improve person tracking accuracy and robustness. State-of-the-art visual detection methods such as the Faster-RCNN with ResNet as feature network provide a good combination and trade-off on performance and speed. Real-time people tracking can be performed following a standard tracking-by-detection paradigm, using the Hungarian assignment for association and a Kalman filter for robust tracking.

We propose to extend our existing through-wall experiments [3, 4], by adopting visual people tracking to generate a rich training dataset for a general device-free RF sensing model. For example, we can design and perform long-term experiments with an NLOS RF sensor deployment and automatic ground truth labeling from the CV tracking system. By collecting a sufficiently large labeled dataset, machine learning can be applied for data-driven RF model parameter optimization.

2.3 Automatic RF parameter learning

We propose a data-driven approach to automatically learn the RF model that can generate occupancy images $\mathbf{x}$ from RF network measurements $\mathbf{d}$, given training dataset $D = (\mathbf{d}_i, \mathbf{x}_i)_{i=1}^N$ obtained from both the visual tracking results and the NLOS experimental datasets. We will investigate the following deep learning methods: (1) deep reinforcement learning (RL), and (2) supervised learning.

For deep RL, we propose to learn a policy that can optimize RF model parameters from a standard initialization. The policy will be updated via Q-learning that iteratively refines the Q-values (state-action pairs) by maximizing rewards, which can be estimated from a loss function reflecting the distance between the current RF tracking images and the ground-truth visual tracking images.

We also propose to directly learn a reconstruction mapping between the RF sensor data and the occupancy images. We can convert the RTI solution, essentially an inverse problem, into a data-driven supervised learning problem. We can also use supervised learning algorithms to estimate model parameters from labeled training data.

2.4 Preliminary experiments

We have setup the RF sensor network and performed experiments on person tracking. Initial results show the effects of different RF model parameters on the quality of the reconstructed motion images. In one experiment, a person walked along a pre-marked path with sixteen RF sensors located in a 8 × 12 feet area as shown in Figure 2. We choose covariance model parameters σ_x^2 and δ as described in § 2.1 to illustrate the model parameter effect on estimated images. Two human motion images are shown in Figure 2 with different parameter values. In general, we found that the images from the KRTI algorithm are less sensitive *w.r.t* the change of only one single parameter. However, results can have significant fluctuations when multiple parameters are changing all together.

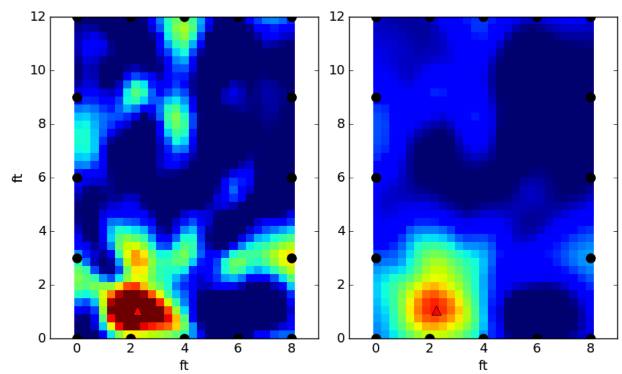

Figure 2: KRTI images of one person in a multipath-rich office environment (pixels with maximum values shown in red triangles are estimated person locations, left image from $\sigma_x^2 = 0.2$, $\delta = 5$; right image from $\sigma_x^2 = 0.1$, $\delta = 20$).

Observe in Figure 2 that large potential improvement can be made for the proposed RF model learning approach, particularly on the selection of model parameters. We expect this improvement can greatly advance device-free tracking in the complex multiple people scenarios with cluttered backgrounds.

3 CONCLUSION AND FUTURE WORK

We proposed to combine RF sensing with visual tracking and use deep neural networks to automatically generate a quality dataset and parameter settings. Early experimental results and performance analysis are shown on the KRTI system. We will further investigate the proposed approach for multiple people device-free tracking.

REFERENCES

[1] Ming-Ching Chang, Yi Wei, Nenghui Song, and Siwei Lyu. 2018. Video Analytics in Smart Transportation for the AIC'18 Challenge. In *CVPR AI City Workshop*.

[2] Jixu Chen, Ming-Ching Chang, Tai-Peng Tian, Ting Yu, and Peter Tu. 2015. Bridging computer vision and social science: A multi-camera vision system for social interaction training analysis. In *ICIP*. 823–826.

[3] Joey Wilson and Neal Patwari. 2010. Radio Tomographic Imaging With Wireless Networks. *IEEE Transactions on Mobile Computing* 9, 5 (May 2010), 621–632.

[4] Yang Zhao, Neal Patwari, Jeff M. Phillips, and Suresh Venkatasubramanian. 2013. Radio Tomographic Imaging and Tracking of Stationary and Moving People via Kernel Distance. In *IPSN*. 229–240.