Human Detection Behind Walls Based on Faster R-CNN Algorithm

Suyun Sun

June 2024

Abstract

Human detection is an important research topic in the field of wireless sensing and computer vision. Most of the researches mainly use cameras and millimetre wave radar, which can achieve fine-grained human localization; however, they cannot solve the problem of wall occlusion. Although methods that use Wi-Fi can penetrate walls, they cannot detect the precise localization of people due to their narrow bandwidth limitation. The above problems limit the application of human detection in real life. This paper proposes a through-wall, multi-person detection system. In this system, we design a multiple-input-multiple-output (MIMO) radar in the 1-2 GHz frequency range. An ultrawideband (UWB) positioning device is attached to the self-developed, through-wall radar system to capture the position information and radar reflection heatmap of people behind a wall, and use the position information as a supervision signal for neural network training. Human detection network based on Faster RCNN aims to learn the human body position information from the radar heat map using the Faster RCNN network to achieve human detection. The experimental results show that the method in this paper can effectively perform human detection for the target human body behind a 24 cm brick wall, and the detection accuracy can reach 92.70

keywords: Faster RCNN; localization; MIMO radar; through wall.

1 Introduction

With the widespread deployment of wireless infrastructure, wireless sensing technology based on wireless signals has garnered extensive attention from researchers. In comparison to traditional camera and infrared-based sensing methods, wireless sensing boasts advantages such as broad coverage and insensitivity to lighting conditions and obstructions. This has led to the emergence of a variety of novel sensing applications, including human intrusion detection, health monitoring, and indoor positioning [1]. An essential application of wireless perception is human detection. Human detection is categorized into active and passive human detection techniques. Active human detection technology necessitates individuals to actively carry various smart devices, which may face limitations in certain scenarios. Therefore, passive detection methods have attracted more attention. In the passive detection process, the target does not need to carry any equipment and can spontaneously undergo 'unconscious continuous perception' of the monitored target. This aids users in achieving ubiquitous perception services at any time and place. From smart homes to smart cities, human detection technology has permeated various aspects of people's lives [2].

There have been numerous research achievements in personnel detection technology based on radar signals. Existing work uses artificial analysis or deep learning methods to extract the true positions of people in space from echo signals. Qian et al. proposed the simultaneous utilization of amplitude and phase information of Channel State Information (CSI) [3], extracting eigenvalues of the amplitude and phase covariance matrices as signal features, and employing support vector machines for state classification. Zhu et al. introduced a novel Robust Passive Motion Detection (R-PMD) scheme [4]. R-PMD incorporates a novel signal filtering technique based on Principal Component Analysis (PCA) and extracts robust histograms of variance probability distributions as environmental detection features. Finally, R-PMD further analyzed the detection performance of different antennas and enhanced the system detection performance by selecting the optimal antenna combination. Zhang et al. proposed an MIMO radar-based method for detecting and localizing human bodies using an improved censored mean-level detector constant false alarm rate (CMLD-CFAR) method to distinguish backgrounds from living bodies with an error within 30 cm without a priori information such as target counting [5]. An essential prerequisite for the aforementioned personnel detection systems is the movement of the detection target, which does not allow for effective detection of static targets. In order to achieve more comprehensive personnel detection, Wu et al. introduced DeMan, a system capable of detecting both stationary and moving personnel [6]. Kocur et al. proposed an ultra-wideband radar-based static personnel localization method, which consists of seven basic stages, including raw radar data preprocessing, background subtraction, target detection, time-of-arrival (TOA) estimation and TOA correlation, wall effect compensation, target localization, and target tracking. The tracking and localization tests were separately conducted for two personnel, and the average accuracy was approximately 10.7 cm [7].

In recent years, the flourishing development of computer vision technology has attracted widespread attention from researchers, particularly after achieving stable breakthroughs in deep learning-based image classification and object detection. This progress has enticed numerous researchers to apply computer vision's image detection technology to personnel detection based on radar signals. Currently, deep learning-based detection algorithms fall into two categories. The first category is two-stage algorithms, such as Fast RCNN. In the detection process, a region proposal box is first generated, followed by the next step of detection for that region proposal box. Due to not being an end-to-end network, the detection speed is suboptimal and challenging to achieve real-time results. However, the region proposal box generated in the first stage facilitates better training in the second stage, leading to improved detection accuracy. The second category is single-stage algorithms, like YOLO [8]. Unlike two-stage algorithms, they directly generate detection structures on the image, employing an end-to-end detection approach. With only one stage, they enhance detection speed, albeit at the cost of slightly reduced detection accuracy compared to two-stage algorithms. Faster R-CNN is widely adopted due to its accuracy in object detection tasks. An improvement over Fast RCNN, Faster RCNN significantly boosts computational speed, striking a good balance between accuracy and speed [9]. Therefore, this study employs Faster RCNN for human detection research using through-wall radar.

2 MIMO Radar System

This section mainly introduces the MIMO radar system used in this article. The system consists of three main components: radar structure, radar imaging, and background removal.

2.1 Radar Structure

The MIMO radar system is comprised of an FMCW signal source, a power amplifier (PA), a mixer, a low-noise amplifier (LNA), an RF switch, an antenna array, and a field-programmable gate array (FPGA), as depicted in Fig. 1(a). The selected signal source outputs an FMCW signal with a bandwidth of 1 GHz (1 2 GHz) and a period of 500 µs. This signal fulfills the requirements for high distance resolution and penetration through brick walls in the through-wall human detection system. Fig. 1(b) illustrates the radar's antenna array, consisting of a two-dimensional planar array with 4 transmitting antennas and 16 receiving antennas. The receiving antennas are horizontally spaced at intervals of 10 centimeters, while the transmitting antennas are vertically spaced at intervals of 20 centimeters. This array, equivalent to a 64-antenna array for the transceiver, allows for the detection of a wider spatial range while obtaining information on the orientation, height, and distance of human targets.

2.2 Radar Imaging

The MIMO radar emits FMCW signals, which pass through walls and are reflected upon the human body. The reflected signal received by the radar is processed by LNA and a mixer to obtain an intermediate frequency signal, which contains the distance information of the target in its frequency. The data we obtain is a three-dimensional matrix of size (Nsamples, NRX, NTX), where NSamples is the number of fast-time sampling points for the intermediate frequency signal, NRx is the number of horizontal receiving antennas, and NTx is the number of vertical transmitting antennas. A one-dimensional Fast Fourier Transform is performed on the NSamples dimension of this numerical matrix to

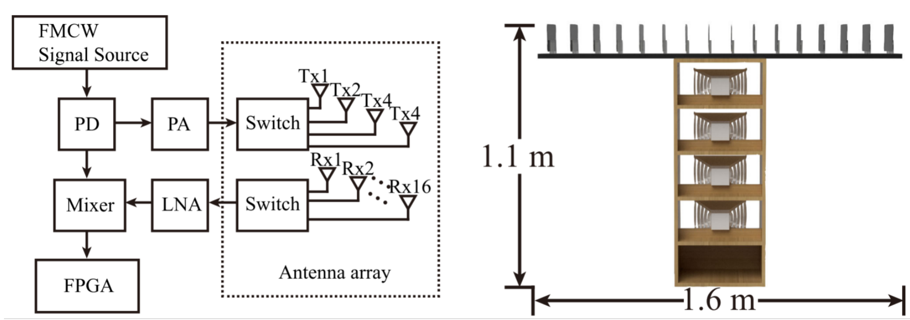

Figure 1: MIMO Radar. (a) Radar structure. (b) Antenna array.

calculate the distance values. Combining with the 2D virtual array element we designed, another 2D Fast Fourier Transform is applied to the NRx and NTx dimensions, obtaining the azimuth and pitch angles of the target space. Finally, we obtain a three-dimensional matrix P of size (range, ρ, θ) representing the space, where each dimension corresponds to distance range, azimuth ρ and pitch angle θ. For the sake of subsequent deep learning, we compress the spatial three-dimensional numerical matrices by averaging them along the pitch angle and azimuth angle directions, resulting in horizontal two-dimensional numerical matrices and vertical two-dimensional numerical matrices. The original horizontal and vertical radar heatmaps are obtained by color mapping these two numerical matrices, as illustrated in Fig. 2.

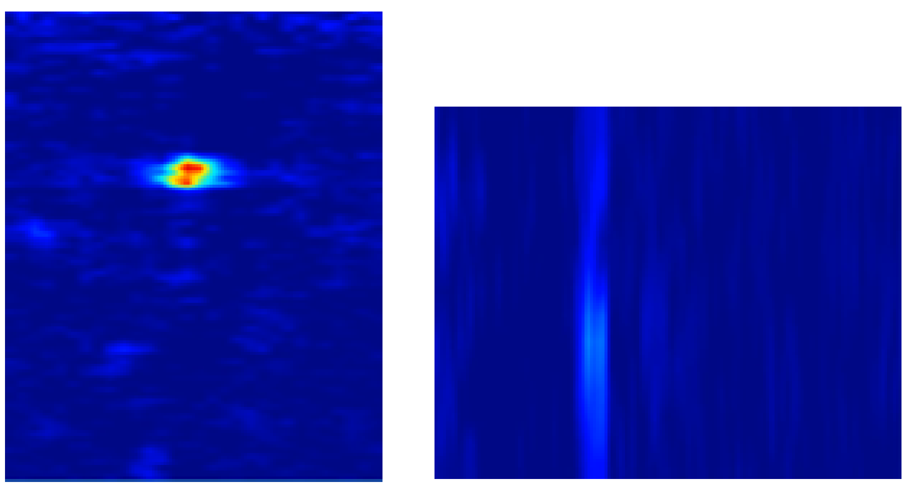

Figure 2: Fig. 2.Original horizontal and vertical radar heatmaps.

3 RPN Loss

3)The loss function of RPN consists of two parts, classification loss and bounding box regression loss. The classification loss is the cross-entropy loss for the foreground and background classification of each anchor frame, as shown in (1). For each anchor frame, the model predicts the probability that it contains an object (foreground); the bounding box regression loss is the difference between the predicted bounding box (i.e., the result of the anchor frame adjustment) and the ground truth. This is typically represented using Smooth L1 loss, which is an L1 loss smoothed around 0 that prevents large errors from affecting training too much, as shown in (2). The term $p_i^* L_{reg}$ means the regression loss is activated only for positive anchors ($p_i^* = 1$) and is disabled otherwise ($p_i^* = 0$).

$$I = \int_a^b f(x)\,dx + \sum_{n=1}^{\infty} \frac{1}{n^2} + \lim_{x \to \infty} \left(1 + \frac{1}{x}\right)^x + \sqrt{\frac{a+b}{c}} \cdot \left(\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y}\right) \tag{1}$$

$$f(x) = \begin{cases} 1, x > 0 \\ 0, x \leq 0 \end{cases} \tag{2}$$

4 Table

Furthermore, we compared our prediction results with the YOLO network [8]. The results of the accuracy comparison for the five sets of sample data are shown in Table I.

Table 1: Comparison of accuracy of five sample groups

Method	Accuracy
Ours	92.70
YOLO	88.67

References

[1] Z. Wang et al., "A survey on human behavior recognition using channel state information," *IEEE Access*, vol. 7, pp. 155986–156024, 2019.

[2] A. Ben Mabrouk and E. Zagrouba, "Abnormal behavior recognition for intelligent video surveillance systems: A review," *Exp. Syst. Appl.*, vol. 91, pp. 480–491, Jan. 2018.

[3] K. Qian, C. Wu, Z. Yang, et al., "Enabling Contactless Detection of Moving Humans with Dynamic Speeds Using CSI," *ACM Transactions on Embedded Computing Systems*, 2018, 17(2): 1-18.

[4] H. Zhu, F. Xiao, L. Sun, X. Xie, P. Yang and R. Wang, "Robust and Passive Motion Detection with COTS WiFi Devices," *Tsinghua Science and Technology (TST)*, 2017, 22 (4): 345-359.

[5] Y. Zhang et al., "A coarse-to-fine detection and localization method for multiple human subjects under through-wall condition using a new telescopic SIMO UWB radar," *Sens. Actuators A, Phys.*, vol. 332, Dec. 2021, Art. no. 113064.

[6] C. Wu, Z. Yang, Z. Zhou, et al., "Non-Invasive Detection of Moving and Stationary Human With WiFi," *IEEE Journal on Selected Areas in Communications*, 2015, 33(11): 2329-2342.

[7] D. Kocur, T. Porteleky, M. Švecová, M. Švingál, and J. Fortes, "A novel signal processing scheme for static person localization using M-sequence UWB radars," *IEEE Sensors J.*, vol. 21, no. 18, pp. 20296–20310, Sep. 2021.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779-788.