Real-Time Indoor 3D Human Imaging Based on MIMO Radar Sensing

Abstract—**Compared to traditional camera-based computer vision and imaging, radio imaging based on wireless sensing does not require lighting and is friendly to privacy. This work proposes a deep learning radio imaging solution to visualize real-time user indoor activities. The proposed solution uses a low-power, MIMO Frequency-Modulated Continuous Wave (FMCW) radar array to capture the reflected signals from human objects, and then constructs 3D human visualization through a serials of data analytics including: $1)$ a data preprocessing mechanism to remove background static reflection, $2)$ a signal processing mechanism to transfer received complex radar signals to a matrix containing spatial information, and $3)$ a deep learning scheme to filter abnormal frames resulted from rough surface of human body. This solution has been extensively evaluated in an indoor research lab. The constructed real-time human images are compared to the camera images captured at the same time. The results show that the proposed radio imaging solution can result in significantly high accuracy.**

I. Introduction

Indoor human imaging is of utmost importance to many intelligent devices or systems. For example, robots need real-time human images to plan and change the route, and smart health system needs human images to recognize their activities for alerts when children or elderly people fall. However, most human imaging solutions are based on cameras, which are associated with privacy concern [1], [2]. Furthermore, lighting is critical to the performance of these solutions. For example, a regular camera does not work well in a dark environment. Hence, radio imaging based on wireless sensing has become a popular research topic. In addition to privacy protection, two more benefits are associated with radio imaging to capture indoor human activities: one is that radio imaging solutions can "see" human objects in dark light conditions, and the other is that radio signals can offer human imaging without line-of-sight requirement, e.g. through a wall.

In this paper, we propose a deep learning radio imaging solution based on MIMO radar sensing. This solution first uses a MIMO FMCW radar to sense environment, then converts the raw signals to 3D human images, which contain spatial information of target in real-time. This solution has the following highlights:

• It designs a framework combining the use of an array of directional beamforming antennas and the FMCW radar sensing to obtain the 3D spatial information of environments, at very low power with an average transmission power of less than -40 dbm/MHz.

• It proposes calibration algorithm to remove the static environment response from raw signals, so that only useful human motion responses can be reserved, which facilitates the visualization of the human activitiy.

• It develops a deep learning algorithm to process raw 3D images. The deep learning model is trained to identify any abnormal frames and remove them to reflect the real activity of a human user in imaging.

• Its design leverages off-shelf low-cost devices and achieves reliable performance in real-time cases.

In the rest of this paper, Section II reviews the literature solutions related to our work. Next, Section III describes the system architecture, including platform, signal processing chains as well as data preprocessing to remove static environmental reflections. Then, Section IV discusses the technical details of combining FMCW radar with directional antenna array to collect and visualize the 3D spatial information based on the signal power at specific spatial voxels. Section V presents a deep learning solution to recognize and remove abnormal frames, followed by Section VI that evaluates the performance of the whole system in real-time cases. This paper is concluded by Section VII.

II. Related work

Many algorithms have been proposed to obtain human activity information based on computer imaging and vision techniques [3]–[7]. Sung et al. used RGB-D depth images to detect and track human motions. Those images with depth information were properly processed to generate the human movement in a 3-D space along the time [3], [6]. Jalal's team proposed a solution that uses scaling invariant features with depth videos to recognize human logging activity [4]. More recently, Microsoft Kinect depth camera was used to collect human motion data because of its abundant APIs. Researchers use trained machine learning models to do image segmentation for Kinect real-time video and obtain the coarse human outlines and motions [5], [7].

Radar sensors and RF devices are usually used for military or wireless communication purpose. However, wireless radio has been recently considered for smart home applications because of the benefits in data confidentiality, and the performance does not depend on lighting conditions. Recent radio imaging research works are either based on FMCW radar [8]–[11] or off-the-shelf devices [12]–[14]. A research group at Massachusetts Institute of Technology (MIT), Adib et al. designed a special MIMO antenna system with FMCW technique based on a software defined radio platform to detect human motion [8] and can capture human movements through a wall [9], [10]. However, their solution can only generate 2D imaging of human. Off-the-shelf devices such as ultrasonic sensor or FMCW radar sensors have also been investigated to recognize human activities [12]–[14]. Avrahami et al. proposed a human activity recognition scheme based on 2D heat maps generated by FMCW sensors [14], while Zhu et al. [13] adopted traditional signal processing algorithms to filter and cluster raw data and thus recognize human activities. Both of them reports an accuracy of over 80% in their outcomes.

III. System Design Overview

To enable real-time indoor human imaging based on wireless radio signals, we propose a solution, called *DeBat* (deep learning bat), that employs MIMO FMCW radar sensing and deep learning to generate 3D human temporospatial imaging, which offers the human activity information along the time in a space. *DeBat* scans a 3D surrounding with FMCW chirps and a beamforming antenna array. While a pair of antennas are able to compute the direct distance between a detected object and the antenna pair using FMCW chirps, the beamforming antenna array is employed to obtain the spatial directions for 3D imaging.

Briefly, *DeBat* works as follows. It first emits FMCW chirps to scan a 3D volume of surroundings. Then the received signals are processed to remove environment background reflections. After that, *DeBat* calculates the reflection powers of any scanned voxel and constructs 3D images. Finally a Deep Neural Network (DNN) based filter algorithm is designed to identify and remove any abnormal eruptive reflection frames, so that the real-time 3D human imaging can be accurately generated.

A. MIMO FMCW Radar Sensing

Our *DeBat* solution introduces an MIMO FMCW sensing design that emits FMCW chirps and collects reflected signals with a beamforming antenna array. This design is based on an off-the-shelf ultra-low power radar sensor called Walabot [15]. The antenna array is laid out on a board with the average power less than -41 dbm/MHz. The beamforming antenna array contains 18 pairs of antennas as shown on Figure 1. The frequency range of FMCW chirps is 3.3 GHz-10 GHz, which is sufficient to detect a direct distance up to 10 meters based on the gradient of FMCW chirp.

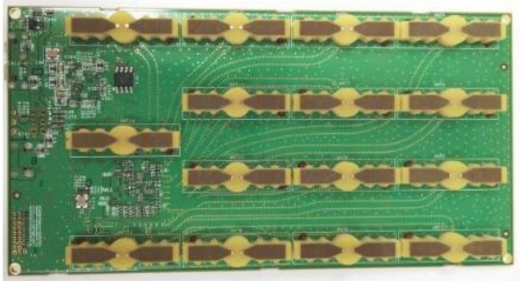


Fig. 1: 2D Antenna Array of Walabot

DeBat beamforms FMCW chirps to scan the angles of ϕ in horizontal direction and θ in vertical direction, as shown on Figure 2 where θ is *elevation* angle to detect the height of human, and ϕ is *wide* angle to capture the width of human. R is the distance between the MIMO FMCW sensor and the human head, which is also the hypotenuse of triangle whose angle is θ. The radar scan is the sectoral pyramid of ϕ, which is the space where the triangle of θ passes in the scan. In our case, θ is from $-45°$ to $45°$ and ϕ is from $-90°$ to $90°$. The distance R can be calculated with FMCW property according to Figure 3 as in Formula (1) below:

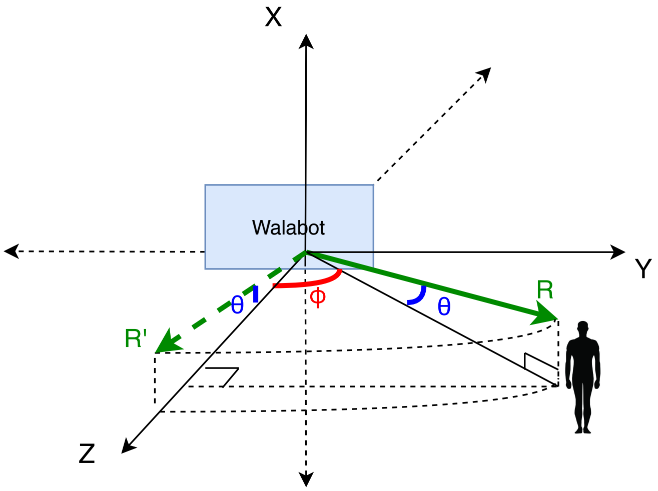


Fig. 2: Scanned 3D Axis of Walabot

$$R = \frac{c|\triangle t|}{2} = \frac{c|\triangle f|}{2(df/dt)} \tag{1}$$

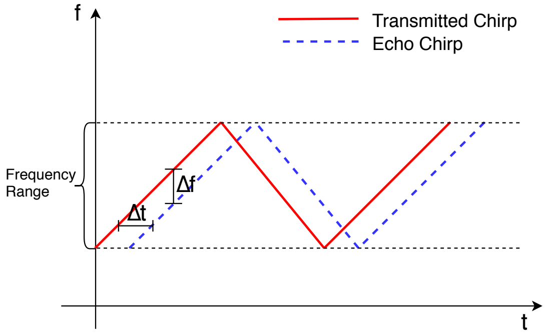


Fig. 3: Frequency Chirp

In the calculation, $\triangle t$ is signal round trip time between the radar sensor and the human object, and $\triangle f$ is frequency difference between the transmitting and receiving signals. df/dt is the slope of a transmitting or echo frequency chirp and c is speed of light. For reasonable simplification, doppler frequency shift effect is not considered.

B. Imaging Flowgraph

The imaging of *DeBat* contains three data processing modules: 1) data collection and calibration, 2) coarse visualization and 3) fine visualization, as shown in Figure 4.

In the first phase, *DeBat* emits FMCW chirps and records static background reflections during initial setup time. Later when *DeBat* starts to scan for human object detection, the signals sensed by its beamforming antenna array are superpositions of both the static background and the human object. The second phase is designed to convert signals to images. Since *DeBat* scans 3D surroundings with parameters R, θ and ϕ, the received signals are processed to represent the energy distribution of every spatial point across different R, θ and ϕ , namely voxel in scanned space. Then *DeBat* subtracts the recorded static background energy to get the energy of the human objects, which is a 3D matrix M with dimension sizes of $(sizeX, sizeY, sizeZ)$, where $sizeX, sizeY, sizeZ$ can be computed with Equation (2), where $range(\cdot)$ is the detection range of parameters, while $res(\cdot)$ is the designated parameters' sampling interval, i.e. scan resolution.

$$\begin{aligned} sizeX &= range(R)/res(R) \\ sizeY &= range(\theta)/res(\theta) \\ sizeZ &= range(\phi)/res(\phi) \end{aligned} \tag{2}$$

In this phase of imaging, because human body acts as an irregular reflector rather than a scatterer, some signals are reflected directly back to antenna array, while others travel through multiple indirect reflective paths, which is the well-known multipath problem in wireless communication. The multipath indirect reflections result in ambiguous and abnormal consequence to the constructed 3D images.

In the third phase, a deep learning scheme is designed to address the multipath indirect reflection problem to achieve the coarse-to-fine visualization. We build a dataset with both regular and multipath reflection images, and then train a Deep Neural Network (DNN) to recognize and remove any multipath reflections in real-time imaging.

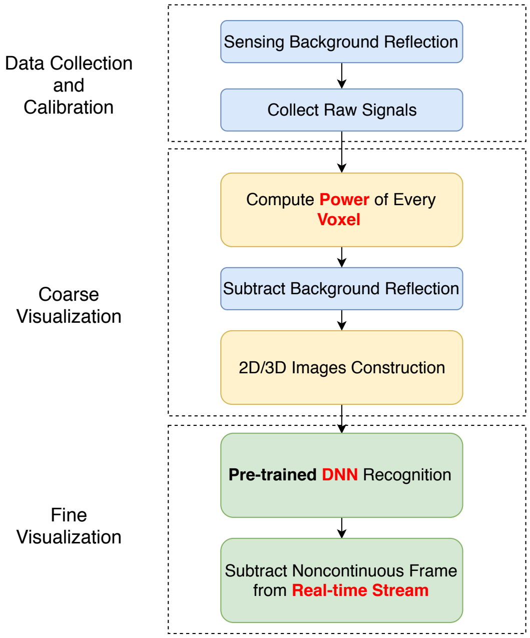
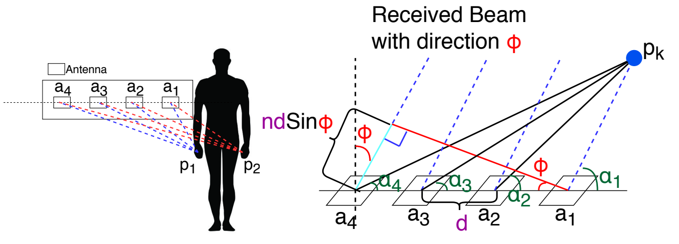

Data Collection
and
Calibration
Sensing Background Reflection
Collect Raw Signals
Coarse
Visualization
Compute Power of Every
Voxel
Subtract Background Reflection
2D/3D Images Construction
Fine
Visualization
Pre-trained DNN Recognition
Subtract Noncontinuous Frame
from Real-time Stream

Fig. 4: Flow Graph of *DeBat*

The following presents the technical details of the *DeBat* system.



(a) Multiple Points **(b)** 2D shape

Fig. 5: 2D Scanning Scenario

IV. Calibration and Visualization

A. Theoretical Formulation

The FMCW signals transmitted and sensed by the antenna array of *DeBat* are complex signals, which can be represented with amplitude and phase as follows:

$$s_t = A_t e^{-j2\pi \frac{r}{\lambda} t} \tag{3}$$

where s_t is the signal received at the moment t . A_t is the amplitude of a signal at time t, r is travel distance of the signal and λ is the wavelength. Since the received signal phase is linear function of the travel distance, $2\pi\frac{r}{\lambda}t$ is the signal phase when it reach the receiving antenna at moment t.

Referring to Equation 3, because the receiver is an antenna array, s_t is rewritten as $s_{n,t}$ to specify the signal is received by which receiving antenna, where n is the n_{th} antenna index. Thus, $s_{n,t}$ refers to a signal received by the antenna n at moment t.

Another parameter needs to be clarified is r. Since human body is a surface rather than a point, it reflects signals from different directions to all antennas, the received signals at moment t of one antenna contains more than one points' reflections. Thus r varies from multiple reflection points. Figure 5a shows a scenario where an antenna array scans a human body. The left hand p_1 reflects to antenna a_1, a_2, a_3, a_4 as blue dot line, and the right hand p_2 reflects to the antenna array as red dot lines. Based on above description, the received signals can be formulated by Equation 4:

$$\begin{aligned} s_{n,t} &= \sum_{k=1}^{K} A_{n,t} e^{-j2\pi \frac{r_{n,k}}{\lambda} t} \\ r_{n,k} &= travel(p_k, a_n) \end{aligned} \tag{4}$$

The donations are p_k is k_{th} points on the detected object, K is the number of points being scanned, $r_{n,k}$ is signal traveling from the reflection p_k to the antenna a_n.

B. Computing Voxel Energy

Energy of Direction: Based on Equations 3 and 4, the problem of computing voxel energy can be formulated as: with known signals $s_{n,t}$ received by antenna a_n at moment t, we need to compute the reflection power of every scanned points. It should be noted that both angles and distance are embedded in the phase of received signal. More specifically, the energy of specific angles ϕ, θ can be derived from antenna array properties, while the energy of distance r can be calculated with FMCW theory. Revisiting Figure 5a and converting the antenna array panel to a plane figure, antennas a_1, a_2, a_3, a_4 receive reflections from p_k, and the incoming direction of beam is ϕ as shown in both Figure 5b and Figure 2. $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are angles between the antenna and p_k, and d is the distance between two antennas. As a result, the energy of direction ϕ can be presented as $P(\phi)$ with Equation 5:

$$P(\phi) = |\sum_{n=1}^{N} s_{n,t} e^{-j2\pi \frac{nd \sin\phi}{\lambda}}| \tag{5}$$

where N is the number of antennas. Because $s_{n,t}$ travels different distance to each antenna, denote the difference with $nd \sin\phi$ as depicting in light blue color. Then, the phase shift on the antenna n is $2\pi \frac{nd \sin\phi}{\lambda}$, where λ is signal wavelength.

Energy of Distance: The travel distance of a signal is related to the direct distance from point p_k to antenna a_n. FMCW measures the reflection depth by calculating the frequency shift between the transmitting and receiving chirps at a moment. Equation 1 shows the FMCW theory. We define v as the slope of frequency chirp versus time, where v is equal to df/dt in Figure 3. The energy of distance r_k can be calculated with phase shift of $s_{t,n}$ as shown below in Equation 6:

$$P(r_k) = |\sum_{n=1}^{N} \sum_{t=1}^{T} s_{n,t} e^{-j2\pi \frac{v r_{n,k}}{c} t}| \tag{6}$$

where r_k is the signal travel distance from point k and T is the duration of each chirp. Because $f = vt$ and $r/c = t_{travel}$, we can easily get the phase shift as $2\pi f t_{travel}$. Thus the energy of r_k is the sum over duration T and all N antennas.

Energy of Voxel: *DeBat* scans 3D surroundings. Refer to Figure 5b where p_k on the same panel of an antenna. However, points in a 3D space need three parameters to locate, either (r, θ, ϕ) in spherical coordinate system or (x, y, z) in cartesian coordinate system. We choose spherical coordinate system because the energy of θ and ϕ can be calculated based on our 2D antenna array. Figure 6 shows how it works.

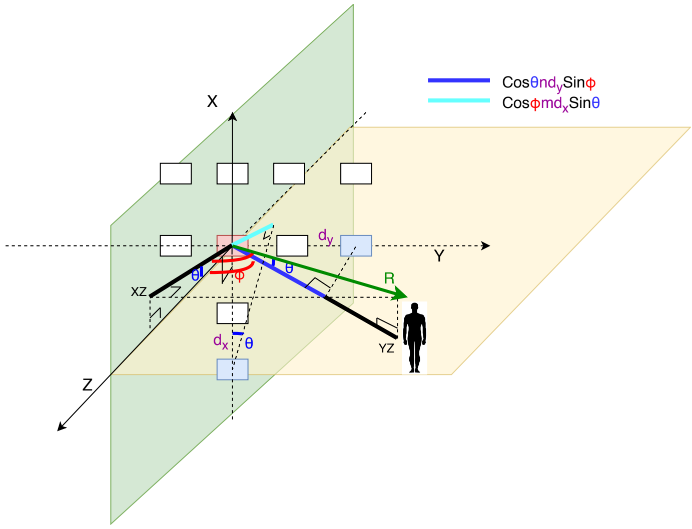


Fig. 6: 3D Voxel Power Description

The 2D antenna array is on $X-Y$ panel, where blocks are antennas. d_x and d_y are distances between two antennas in 2D space. $Y-Z$ panel is the dimension drawn in Figure 5b, and d_y, ϕ are the d, ϕ in Equation 5, while θ is the elevation angle from the $Y-Z$ panel to R. In 3D space, R is converted to the $Y-Z$ panel as YZ with $\cos\theta$, and it is converted to $X-Z$ panel as XZ with $\cos\phi$, where ϕ is wide angle from the YZ panel to Z axis. Thus the distance change on the $Y-Z$ panel for each antenna is $\cos\theta * nd_y * \sin\phi$ as blue line, and that change on the $X-Z$ panel is $\cos\phi * md_x * \sin\theta$ as light blue line shows. Since distance change represents phase shift of a signal, we can calculate energy of any voxel with Equation 7. $s_{n,m,t}$ is the signal received by the antenna n and transmitted by the antenna m at time t.

$$P(r_k, \theta, \phi) = \\ |\sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{t=1}^{T} s_{n,m,t} e^{-j2\pi \frac{v r_k}{c}} e^{j\frac{2\pi}{\lambda} \cos\theta (nd_y \sin\phi + md_x \cos\phi)}| \tag{7}$$

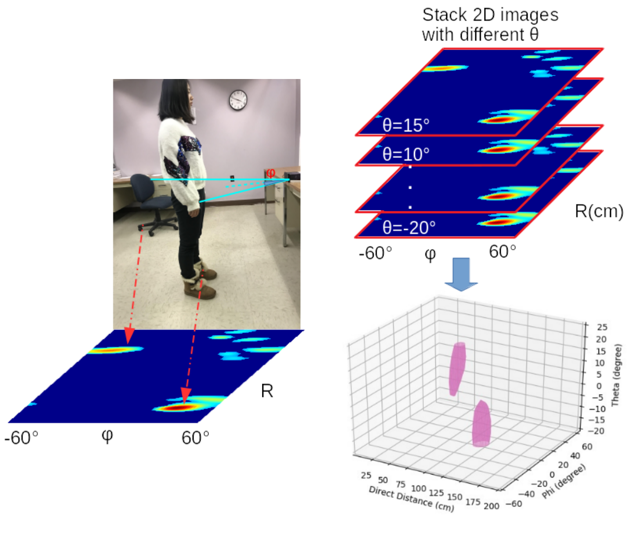
C. 3D Imaging

Removing Background Reflection: To get rid of the environment reflections such as desks or walls, *DeBat* starts a sensing process before capturing humans, which is called calibration. Since the background reflection is static and the reflection energy is fixed, calibration is performed as sensing, calculating and recording the background reflection energy of any voxel. Later when *DeBat* starts human imaging, it subtracts the static background reflection energy from the real-time reflection energy.

Constructing 2D/3D Images: Once *DeBat* calculates the energy of every voxel and removes background reflection power, it generates a 3D matrix M with the dimension of $(sizeX, sizeY, sizeZ)$, where $sizeX, sizeY, sizeZ$ can be referred from Equation 2.

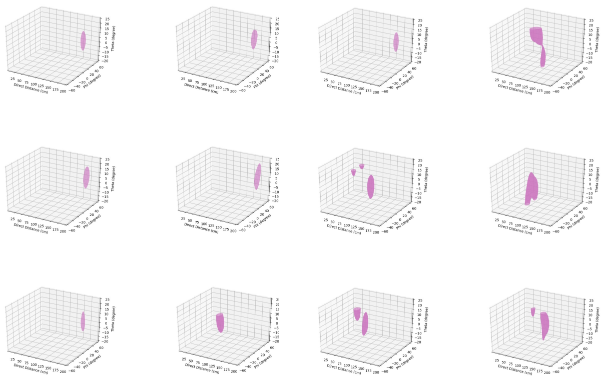
A 2D image is related to either (R, θ), (R, ϕ) or (θ, ϕ). We choose to construct 2D image with distance and wide angle (R, ϕ). At first, we find the highest energy in M. Suppose the highest reflection energy is from the point (R_a, θ_b, ϕ_c), and $M(R_a, \theta_b, \phi_c)$ is the highest value in M. Then $M(R, \theta_b, \phi)$ is a 2D array because parameter θ is fixed as θ_b. Thus we can generate a 2D heatmap image based on $M(R, \theta_b, \phi)$, where the color shows reflection energy intensity, the darker color for the higher energy at (R, θ_b, ϕ). Figure 7a shows a 2D image scenario and its corresponding heatmap generated by *DeBat*.

As can be observed from Figure 7a, the range of ϕ is from $-60°$ to $60°$, where ϕ is the angle from dash blue line to human. In this case where the dash blue line is the base line in the middle of the sensor, ϕ is the wide angle from the base line to the object. While 2D image only depicts the highest power layer of fixed θ, a 3D image shows more information about object width, height and location. Figure 7b shows how to construct 3D images from 2D images. *DeBat* uses marching cubes algorithm [16] to generate vertices and faces of the stacked images, then it uses a normalized filter to remove low



(a) 2D Heatmap **(b)** Stack to 3D Image

Fig. 7: 2D/3D Image Capturing



(a) Regular Dataset **(b)** Ambiguours Dataset

Fig. 8: Dataset Overview

power points. From the 3D images, it is very clear to see that the human object is at a shorter direct distance to radar, and the height of human is greater than the chairs.

V. Abnormality Detection and Removal

One challenge of real-time human radio imaging is signal deviation. Since human body is not a plane surface, especially when a human moves, the surface of body is extremely deformed. As a result, while the antenna array transmits signals and scans human body, only some signals are directly reflected back toward the antennas. Other signals may be deviated through multiple indirect paths back to the receiver, which makes the system "misunderstand" the real distance and angle of the object. This scenario can be clearly observed in Figure 9:

In this case, the distance between the human chest and legs is not quite large. However, signals reflected by human legs deviate their incoming path, and as a result the receive antenna gets signals from $r_3\theta_3\phi_3$, where $r_3 = d_1 + d_2$. That antenna ”misunderstands” the human leg position with wrong distance and angles, and it results in a deformed 3D shape. To address this issue, we design a Deep Neural Network to decide whether the current 3D figure is deformed or not, and remove any abnormal images from the image stream.

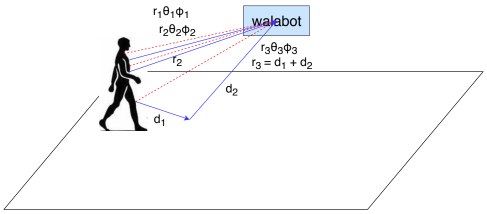


Fig. 9: Signal Deviation

DNN Recognization: We use the transfer learning technique to solve this problem more efficiently. Because this is an image processing problem, the proper Deep Neural Network should have convolutional layers to reduce the amount of possible parameters and calculation. Based on that, we choose **resNet18** to classify our 3D images. In this work, we 1) collect regular and ambiguous images to build a training dataset, 2) change the network structure of resNet18 to design a faster DNN, and 3) load the pre-trained DNN parameters to remove bad frames.

Building Dataset: We collect training dataset from real human activities, where one person walks around in the lab. We construct 3D images and concatenate them into 3D videos. Then we label them manually into two categories: regular frames and ambiguous frames. Figure 8 shows samples of the dataset. Figure 8a shows the regular 3D reflection energy and Figure 8b displays ambiguous images. As can be observed from the dataset, the regular frames show human 3D position very clear, while the ambiguous frames always ”misunderstand” locations of some parts of the human body.

Redesigning resNet18 DNN: The first step to apply transfer learning is to change the last Fully Connect (FC) layer. The last FC layer of a normal resNet18 is of $(512, 1000)$, which means the FC layer is 512 input features and 1000 output features. The 1000 $out\ feature$ usually feed into $softmax$ functions to be classified into 1000 categories. In our design, we only have 2 categories: regular and ambiguous. Thus we change the dimension of last FC layer to $(512, 2)$. The second revision of the original resNet18 is on the pooling layer before the FC layer. Resnet18 uses an Average Pooling layer to compress features to 512. However, Average Pooling sometimes cannot extract good features because it takes all values for an average. Since our dataset images have strong edges, **Max Pooling** is better to extract the most important or extreme features.

Loading DNN: In the *DeBat* system, the trained parameters of DNN are loaded into the main process. Every frame generated is fed into the DNN for classification. It is recognized and labeled by the DNN with either "regular" or "bad" labels. If a frame is recognized as "bad", the main process removes it and reuse the previous frame as the current frame for visualization . Since the loading and detection process are extremely fast, *DeBat* can achieve real-time imaging.

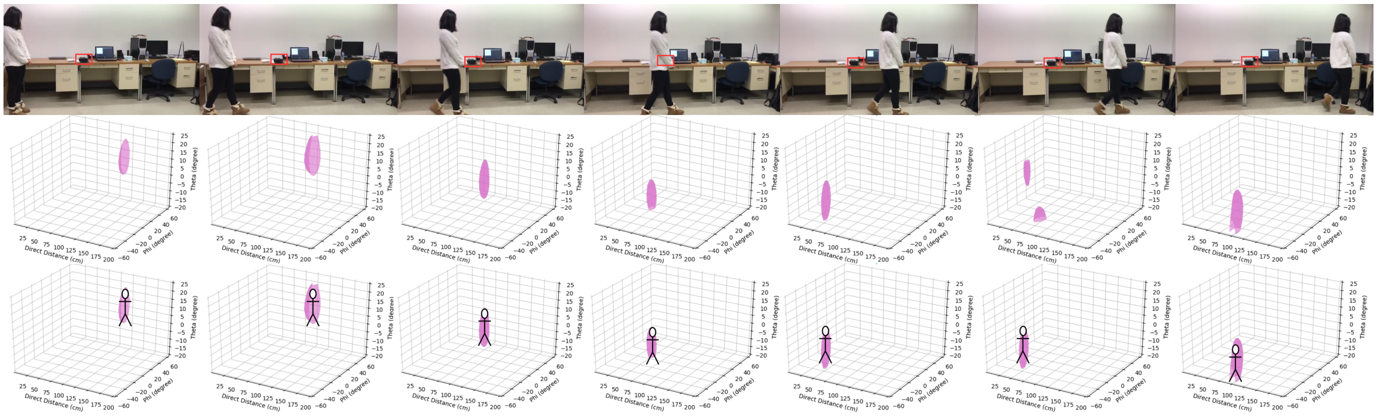


Fig. 10: Human Real-time Image Capturing

VI. Performance

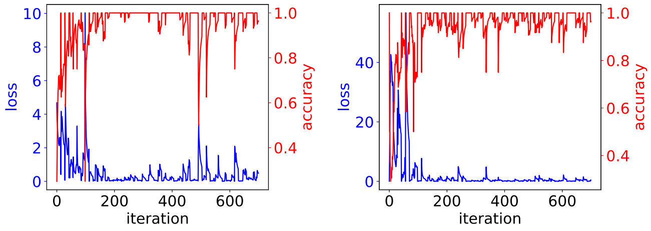
In this section, we evaluate the performance of our 3D imaging solution as well as its DNN classification. The 3D imaging parameters are set based on our lab size, with the direct distance from 0 to 200 centimeters, the wide angle Phi(ϕ) from $-60°$ to $60°$, and the elevation angle Theta(θ) from $-20°$ to $25°$.

Hyper-parameters of DNN: The DNN is trained with mini-batch strategy to make it converge more smoothly. We use the cross entropy as loss function as shown in Equation 8 where x is output of DNN, whose dimension is $(minibatchsize, 2)$, and $label$ is the labels for one mini-batch data with dimension $(minibatchsize, 1)$. We use a SGD optimizer to update parameters with $learning\ rate = 0.01$ and $momentum = 0.9$, and a $lr\ schedular$ is applied to adjust the learning rate with $stepsize = 7$ and $gamma = 0.1$. Then we collect the running loss and accuracy of each iteration and plot in Figure 11. Note that the running loss and accuracy are cleared after each epoch.

$$loss(x, label) = -log(\frac{e^{x[label]}}{\sum_j e^{x[j]}}) \tag{8}$$

Figure 11a shows the original performance of resnet18 and Figure 11b shows the results of our revised DNN. It is clearly to observe that our DNN converges faster and has less vibrations compared to the original resnet18.

Imaging of DeBat: We conduct the evaluation of *DeBat* imaging in a scenario where a human object moves around in the lab. The results are shown in Figure 10. The first row records real human motions, the second row is the results before filtering abnormal frames, and the third row is a final result of the imaging. At the very beginning, human is standing on the right of radar with a wide angle of 60°, where the cube in rows 2 and 3 stand around $\phi = 60°$ and $R = 110cm$. With human moving closer to the radar in frames $1-3$, the captured images show that the wide angle ϕ and the direct distance R are decreasing gradually. While the human moves away from the radar, the wide angle ϕ and the direct distance R increase. During this time, a frame is detected by the DNN as "bad", and the previous frame is reused as the current one in the visualization.



(a) Average Pooling Result

(b) Max Pooling Result

Fig. 11: 2D/3D Image Capturing

VII. Conclusion

In this paper, we propose a real-time 3D indoor human imaging solution based on MIMO FMCW radar sensing. This solution not only localizes human position precisely in a 3D space, but also protects the privacy. This solution also introduces a deep learning model to remove any abnormal imaging resulted by multipath reflection problem. Our future work will focus on recognizing human activities based on the radio imaging generated by this solution.

References

[1] Daphne Townsend, Frank Knoefel, and Rafik Goubran. Privacy versus autonomy: a tradeoff model for smart home monitoring technologies. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 4749–4752. IEEE, 2011.

[2] J Sathish Kumar and Dhiren R Patel. A survey on internet of things: Security and privacy issues. *International Journal of Computer Applications*, 90(11), 2014.

[3] Jaeyong Sung, Colin Ponce, Bart Selman, and Ashutosh Saxena. Human activity detection from rgbd images. *plan, activity, and intent recognition*, 64, 2011.

[4] Ahmad Jalal, Md Zia Uddin, and T-S Kim. Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home. *IEEE Transactions on Consumer Electronics*, 58(3), 2012.

[5] Lu Xia, Chia-Chih Chen, and Jake K Aggarwal. Human detection using depth information by kinect. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, pages 15–22. IEEE, 2011.

[6] Matteo Munaro, Christopher Lewis, David Chambers, Paul Hvass, and Emanuele Menegatti. Rgb-d human detection and tracking for industrial environments. In *Intelligent Autonomous Systems 13*, pages 1655–1668. Springer, 2016.

[7] Xiaojun Chang, Zhigang Ma, Ming Lin, Yi Yang, and Alexander G Hauptmann. Feature interaction augmented sparse learning for fast kinect motion detection. *IEEE transactions on image processing*, 26(8):3911–3920, 2017.

[8] Fadel Adib and Dina Katabi. *See through walls with WiFi!*, volume 43. ACM, 2013.

[9] Fadel Adib, Zachary Kabelac, Dina Katabi, and Robert C Miller. 3d tracking via body radio reflections. In *NSDI*, volume 14, pages 317–329, 2014.

[10] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. Capturing the human figure through a wall. *ACM Transactions on Graphics (TOG)*, 34(6):219, 2015.

[11] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. Through-wall human pose estimation using radio signals. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7356–7365, 2018.

[12] Shangyue Zhu, Hanqing Guo, Junhong Xu, and Shaoen Wu. Distance based user localization and tracking with mechanical ultrasonic beamforming. In *2018 International Conference on Computing, Networking and Communications (ICNC)*, pages 827–831. IEEE, 2018.

[13] Shangyue Zhu, Junhong Xu, Hanqing Guo, Qiwei Liu, Shaoen Wu, and Honggang Wang. Indoor human activity recognition based on ambient radar with signal processing and machine learning. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2018.

[14] Daniel Avrahami, Mitesh Patel, Yusuke Yamaura, and Sven Kratz. Below the surface: Unobtrusive activity recognition for work surfaces using rf-radar sensing. In *23rd International Conference on Intelligent User Interfaces*, pages 439–451. ACM, 2018.

[15] Walabot. https://walabot.com/.

[16] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *ACM siggraph computer graphics*, volume 21, pages 163–169. ACM, 1987.