# mmEye: Super-Resolution Millimeter Wave Imaging

Feng Zhang, *Member, IEEE*, Chenshu Wu, *Member, IEEE*, Beibei Wang, *Senior Member, IEEE*, and K. J. Ray Liu, *Fellow, IEEE*

***Abstract*—RF imaging is a dream that has been pursued for years yet not achieved in the evolving wireless sensing. The existing solutions on WiFi bands, however, either require specialized hardware with large antenna arrays or suffer from poor resolution due to fundamental limits in bandwidth, the number of antennas, and the carrier frequency of 2.4 GHz/5 GHz WiFi. In this article, we observe a new opportunity in the increasingly popular 60-GHz WiFi, which overcomes such limits. We present mmEye, a super-resolution imaging system toward a millimeter-wave camera by reusing a single commodity 60-GHz WiFi radios. The key challenge arises from the extremely small aperture (antenna size), e.g., < 2 cm, which physically limits the spatial resolution. mmEye's core contribution is a super-resolution imaging algorithm that breaks the resolution limits by leveraging all available information at both the transmitter and receiver sides. Based on the MUSIC algorithm, we devise a novel technique of joint transmitter smoothing, which jointly uses the transmit and receive arrays to boost the spatial resolution while not sacrificing the aperture of the antenna array. Built upon this core, we design and implement a functional system on commodity 60-GHz WiFi chipsets. We evaluate mmEye on different persons and objects under various settings. Results show that it achieves a median silhouette (shape) difference of 27.2% and a median boundary keypoint precision of 7.6 cm, and it can image a person even through a thin drywall. The visual results show that the imaging quality is close to that of commercial products like Kinect, making for the first-time super-resolution imaging available on the commodity 60-GHz WiFi devices.**

***Index Terms*—60-GHz WiFi, millimeter wave (mmWave), multiple signal classification (MUSIC) algorithm, super-resolution imaging.**

## I. INTRODUCTION

RF IMAGING is a long standing, challenging problem in the evolving community of wireless sensing. Given the ubiquity of WiFi, it is of particular interest to enable imaging using WiFi reflection signals, i.e., creating images of humans and objects without attaching any radio source to the target. It would enable pervasive human and object sensing with rich contexts (not only the location but also silhouette/shape, size, pose, etc.) for new applications, such as interactive gaming, pose estimation, exercise assessment, human recognition, etc., all in a privacy preserving, lightweight, and cost-effective way.

Manuscript received May 28, 2020; revised September 7, 2020 and October 6, 2020; accepted November 6, 2020. Date of publication November 16, 2020; date of current version April 7, 2021. *(Corresponding author: Beibei Wang.)*

Feng Zhang was with the Department of Research and Development, Origin Wireless Inc., Greenbelt, MD 20770 USA. He is now with the Amazon.com Inc., Seattle, WA, USA (e-mail: fzhang15@umd.edu).

Chenshu Wu, Beibei Wang, and K. J. Ray Liu are with the Department of ECE, University of Maryland, College Park, MD 20742 USA, and also with the Department of Research and Development, Origin Wireless Inc., Greenbelt, MD 20770 USA (e-mail: cswu@umd.edu; bebewang@umd.edu; kjrliu@umd.edu).

Digital Object Identifier 10.1109/JIOT.2020.3037836

Despite of the advances in reusing WiFi devices for wireless sensing [1], [2], such as activity sensing [3], gesture recognition [4], vital sign monitoring [5], etc., RF imaging faces significant challenges and remains unsolved. Prior works related to RF imaging on WiFi bands can track human motion and activities [6], [7], map large obstacles [8], and detect malicious objects [9]. But they require specialized hardware with large antennas unavailable on commodity radios or suffer from poor imaging quality. Wision [10] explores the feasibility of using 2.4-GHz WiFi radios with multiple antennas for object imaging, but the resolution is inherently limited by WiFi signals at 2.4 GHz/5 GHz. Recent works strive to estimate human figures (i.e., skeletons, and poses) [11], [12] with a neural network trained by video, but again require multiple specialized FMCW radars with relatively large arrays. Generally, the imaging capability of 2.4 GHz/5 GHz WiFi is fundamentally limited by narrow bandwidth, small antenna number, and large wavelength. While existing millimeter-wave (mmWave) systems can offer high precision imaging [13]–[15] with large lens radars and dedicated circuits, they are all specialized radars and not suitable for ubiquitous applications. [16] took the first step in RF imaging using 60-GHz networking radios. However, it focuses on imaging objects only and requires to precisely move the receiver, without fully exploring advantages from 60-GHz radios.

Recently, two new opportunities have arisen in the design of WiFi imaging systems as follows.

1) 60-GHz networking radios are emerging as 60-GHz WiFi (e.g., 802.11ad/ay [17]), which is already available in commercial routers [18] and is being integrated in smartphones and in cars [19]. Compared to 2.4 GHz/5 GHz bands, 60-GHz radios offer several distinct advantages: millimeter-wavelength on a high-frequency band, highly directional links enabled by a large phased array (e.g., 6×6 elements), and usually large bandwidth that underpins high ranging resolution (e.g., < 5 cm).
2) In addition to networking, commodity 60-GHz radios are going to support a dual role of radar-like sensing [20], with merely one extra antenna array attached to the chipset and without any circuit changes. With this, the 60-GHz radio, under the radar mode, can transmit and receive on a single networking device and capture the precise channel response for precise sensing and imaging.

In this article, we leverage the foregoing opportunities and present mmEye, a super-resolution RF imaging system toward a mmWave "camera" using a single commodity 60-GHz WiFi device. mmEye leverages the 60-GHz networking radio with its unexplored radar sensing capability. It can image both humans, either moving or stationary with different poses, and objects of various shapes, sizes, and materials. It can even image through a thin drywall, despite the high attenuation of 60-GHz signals.

Even with the radar operations, however, enabling imaging on commodity 60-GHz WiFi radio entails great challenges. For example, purposed for networking, the device is not calibrated as well as and thus not as stable as conventional radar, resulting in fluctuating signal responses. Additionally, reflection signals may be frequently missed due to the inherent high attenuation and directionality of 60-GHz signals. The biggest challenge, however, is to achieve high imaging accuracy with the compact 60-GHz array with a small aperture, a key factor that determines the imaging resolution. In general, the imaging resolution of a radar system is defined by resolution $\propto$ wavelength $\times$ distance/aperture, which is about 28 cm at 1m distance for our experimental device with an antenna array size of 1.8 cm $\times$ 1.8 cm. Prior works attempt to extend the effective aperture by synthetic array radar (SAR), which, however, requires receiver movements and is highly sensitive to the moving trajectory tracking error. The impact of the trajectory error becomes particularly noticeable when the error is greater than the wavelength, which is likely to occur for 60-GHz signals with a wavelength of 5 mm.

Differently, mmEye devises a super-resolution algorithm to break through the resolution limited by the physical aperture and enable precise imaging on commodity 60-GHz radio. The proposed algorithm roots in multiple signal classification (MUSIC) [21], one of the most widely used spatial spectrum estimation techniques, and achieves super-resolution through a novel joint transmitter smoothing (JTS) technique.

First, instead of using the on-chip analog beamforming, we perform digital beamforming on the received signals, which yields a much higher spatial resolution. The analog beamforming built-in the radio usually only provides coarse beam resolution (e.g., 3-dB beamwidth of 15° for our device). We boost the spatial resolution by using the MUSIC algorithm. We perform MUSIC over each spherical surface of different azimuths and elevations at every specific range, estimating the spatial spectrum of the signals reflected off the target at that range. The spatial spectrum, along with the accurate range information offered by the 60-GHz radio, will together reconstruct an image of the target. MUSIC can be used for imaging since the signals are sparse on each spherical surface. However, it is not directly applicable since it suffers from the rank deficiency issue, i.e., the rank of the correlation matrix of the signal space is smaller than the number of actual incoming signals.

To overcome the rank deficiency problem, we employ spatial smoothing (SS) in 2-D space [22], a technique to split the receive array into several overlapped subarrays that reuse the same steering vectors. By adding one more subarray, it is approved that the rank of the correlation matrix of signals increases by 1 with probability 1 [23]. In addition to the spatial subarrays, mmEye utilizes the time diversity of consecutive measurements to estimate the correlation matrix. The synthesized spatial and temporal smoothing effectively solves the rank deficiency issue and significantly reduces the variance of the spatial spectrum estimation by MUSIC.

SS on the receive array, however, further reduces the small antenna array size, i.e., the effective aperture, thereby degrading the imaging precision. To increase the rank without loss in aperture, we propose a novel 2-D SS that jointly reuses the transmit array and the receive array, termed as JTS. Specifically, rather than dividing the receive array into subarrays, we reuse the entire receive array for each individual transmit antenna as a subarray. Given our case of 32 transmitter elements, we immediately obtain 32 subarrays, offering a guaranteed rank of 32, which is adequate for the sparse reflection signals, while retaining the scarce aperture unimpaired. Since the subarray size is as big as the whole receive array, the imaging resolution is maximized. Besides the improvement on the spatial resolution, the JTS scheme also alleviates the well known specularity problem for RF imaging,[1] that the signals reflected off the target may not be captured due to the inherent high attenuation and directionality of the mmWave signals, by utilizing the transmit diversity.

Based on the super-resolution algorithm, we design and implement a functional system of mmEye with additional components on the background and noise cancelation (BANC) and adaptive target detection, etc. We prototype mmEye on commodity 60-GHz networking chipset attached with an additional array and perform experiments with different subjects, locations, and postures. The results demonstrate that mmEye achieves accurate imaging results visually close to a Kinect depth sensor [25], as shown in Fig. 1, with a median silhouette (shape) difference of 27.2% and a median boundary keypoint precision of 7.6 cm at the range of 1 m. With the encouraging performance on a single networking device, we believe mmEye takes an important step toward a ubiquitous mmWave camera and the first step toward dual roles of networking and radar sensing for commodity 60-GHz WiFi radios.

In summary, the main contributions of this article are listed as follows.

1) mmEye leverages the sparsity of the reflection signals off the target at individual ranges and applies MUSIC with a novel JTS, which exploits the Tx diversity[2] to boost the imaging resolution and robustness.
2) Various signal processing techniques, including background and noise cancellation, target detection, are utilized to combat different kinds of defects inherent in the RF system.
3) We prototype mmEye using a Qualcomm 802.11ad chipset and conduct extensive real-world experiments. It shows that mmEye achieves comparable imaging with commercial products like Kinect using a single 60-GHz networking device in a much smaller size, underlying pervasive imaging for various applications, such as VR gaming, pose estimation, etc.

[1]Note that the specularity problem for mmWave signals is much less severe when compared with 2.4 GHz/5 GHz signals [7], [12] since specular reflection happens when the surface roughness of an object is smaller than the wavelength according to Fresnel's Law [24].

[2]The 60-GHz WiFi chip is equipped with an antenna array at the Tx side as well, which is different from the most of traditional Radars.

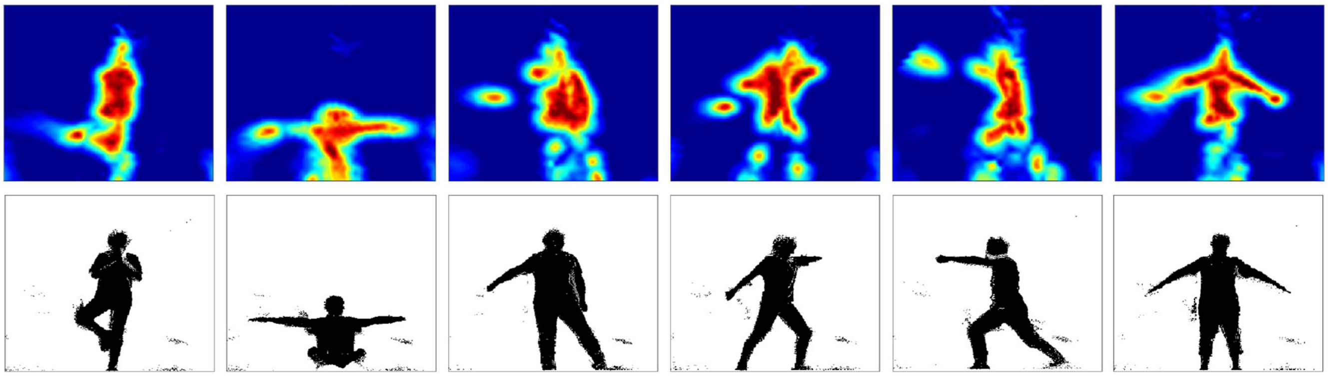


Fig. 1. Example imaging results produced by mmEye (top) compared with Kinect (bottom). The subjects are about 1 m away from the device.

We believe that the design of mmEye will also benefit and inspire future research on sensing using mmWave radios.

The remaining part of this article proceeds as follows. First, we present a primer on 60-GHz WiFi in Section II. Then we introduce the core super-resolution algorithm in Section III, followed with system design in Section IV. We evaluate mmEye in Section V and discuss future works in Section VI. We review the literature in Section VII and conclude this article in Section VIII.

## II. Reusing 60-GHz WiFi as Radar

### A. 60-GHz WiFi

60-GHz WiFi technology, also known as WiGig, with the established IE 802.11ad/ay standards and low-cost commercial chipsets [17], [18], is becoming the mainstream in wireless devices to enable high rate networking and rich user experience. Different from the 2.4 GHz/5 GHz WiFi that faces fundamental limitations in imaging, 60-GHz WiFi offers unique advantages for RF imaging. While the common 2.4 GHz and 5-GHz WiFi devices have only 2 to 3 antennas and 20MHz/40MHz bandwidths, 60-GHz WiFi radios offer many-antenna phased arrays in compact forms and large bandwidths centered at high-frequency band of 60 GHz. These properties translate into several superior features for sensing as follows.E

1) The large phased array enables highly directional beamforming with good spatial resolution.
2) The large bandwidth offers high ranging accuracy.
3) The high carrier frequency leads to more predictable signal propagation that is immune to the multipath effects, a huge challenge for 2.4 GHz/5 GHz WiFi.
4) The carrier wavelength is 5mm, over 10× shorter tha 5-GHz WiFi. This means the required antenna aperture can be 10× smaller to achieve the same imaging resolution.

Additionally, we observe two trends that further promote 60-GHz WiFi as an attractive solution for ubiquitous sensing and imaging as follows.

1) 60-GHz networking chipsets is going to support an additional role of radar-like processing, without hardware changes except for merely one extra antenna array for full-duplex radios, allowing rapid and precise phase measurement with synchronized, co-located transmitter (Tx) and receiver (Rx).

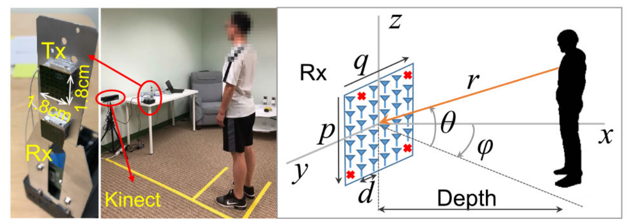



Fig. 2. Device setup and coordinate system. $\theta$ and $\phi$ denote the elevation and azimuth, respectively, and $r$ denotes the range from the device to the reflector. The antenna array contains 32 elements in a $6 \times 6$ layout, with four missing locations marked by red crosses.

2) The commercial 60-GHz WiFi, already used in consumer-grade routers, is becoming relatively inexpensive with increasing market adoption and will soon be available on mobile devices.

Pioneer works have explored 60-GHz radios for tracking and sensing [16], [26], [27]. However, they mainly utilize amplitude information and employ mechanical horn antennas to emulate beam steering. Great potentials in the steerable phased arrays and the dual radar mode of 60-GHz WiFi remains largely underexploited.

### B. 60-GHz WiFi Radar

As shown in Fig. 2, we use commodity Qualcomm 802.11ad chipsets. To enable full-duplex radar operation, an extra array is attached to the chipset to form co-located and synchronized Tx and Rx.[3] The Tx transmits pulses of a known sequence, which, after reflection on surrounding targets, are received and correlated on the Rx side to estimate channel impulse response (CIR) with precise amplitude and phase information.

Suppose $N$ elements in the Tx array and $M$ elements in the Rx array. The CIR between the $n$th transmit antenna and the $m$th receive antenna $h_{m,n}(\tau)$ at time slot $t$ can be expressed as

$$h_{m,n}(\tau, t) = \sum_{l=0}^{L-1} a_{m,n}^{l}(t)\delta(\tau - \tau_l(t)) \tag{1}$$

where $\delta(\cdot)$ is the Delta function, $L$ is the number of the total CIR taps, and $a_{m,n}^{l}$ and $\tau_l$ denote the complex amplitude and

[3]In practice, the dual networking and radar role can be achieved by rapid switching in time., since the radar sensing only requires minimal time. Under the networking mode, the extra array simply provides additional spatial diversity.

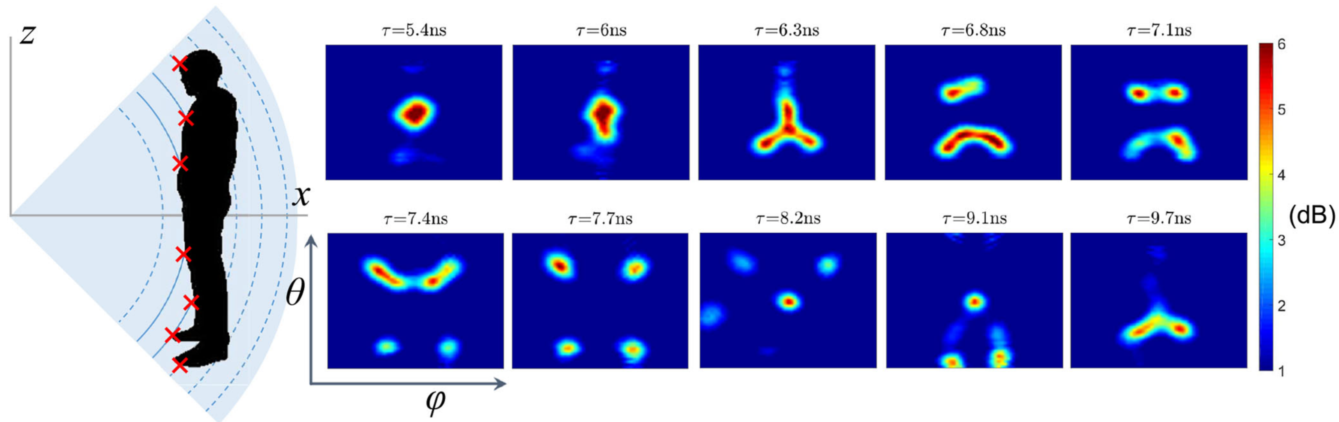



Fig. 3. Spatial spectrum for different propagation delay $\tau$, which demonstrate the sparsity of the reflected signals off the target at individual range/delay.

the propagation delay of the $l$th tap, respectively. To simplify the notations in the following, we omit the dependence on the measurement time $t$ if not mentioned. The time resolution $\Delta\tau$ of the measured CIR is determined by the bandwidth $B$ of the transmitted signal, i.e., $\Delta\tau = 1/B$. Then, $\tau_l$ can be expressed as $\tau_l = \tau_0 + (l-1)\Delta\tau$, where $\tau_0$ denotes the Time of Arrival (ToA) of the first tap. At each time slot, mmEye captures $M \times N \times L$ complex values, i.e., $h_{m,n}(\tau_l)$, where $m = 1, \ldots, M$, $n = 1, \ldots, N$, and $l = 0, \ldots, L-1$. The 3-D information of the target being imaged can be thus inferred from these measurements.

Specifically, our experimental device has 32 elements assembled in a 6×6 layout.[4] for both Tx and Rx (i.e., $N = M = 32$) and operates at 60-GHz center frequency with a 3.52-GHz bandwidth. The measured CIR thus offers a propagation delay resolution of $\Delta\tau = 0.28$ ns, corresponding to a range resolution of 4.26 cm. When the device is set to radar mode, each CIR $h_{m,n}(\tau)$ is measured in a sequential way as follows: the $n$th transmit antenna transmits an impulse while other transmit antennas keep silent, and only the $m$th receive antenna records the corresponding CIR at the same time.[5] The above channel sounding process would repeat $32 \times 32 = 1024$ times in total for a complete CIR recording.

## III. Super-Resolution Imaging

RF imaging leverages the observation that the energy distribution of the reflected RF signals over the space would sketch the silhouette of a target. mmEye tries to reconstruct the contour of the target based on the estimation of the Angle of Arrival (AoA) and ToA of each signal reflected off the surface of the target. As mentioned above, however, the spatial resolution is greatly limited due to the small effective aperture of the receive antenna array. For example, the on-chip analog conventional beamforming (CBF) only provides a 3-dB beamwidth of 15°, which is inadequate to image a target, especially when the target is far away to the device.

To boost the spatial resolution, mmEye performs digital beamforming on the received CIR as opposed to the on-chip analog beamforming, which achieves higher resolution in distinguishing the signals radiated by nearby parts of the target. Noticing that the CBF and the well-known minimum variance distortionless response (MVDR) beamforming (also known as Capon beamformer) [28] both produce poor precision, we devise in this work a super-resolution algorithm based on MUSIC [21], one of the most widely used algorithms for AoA estimation.

### A. Imaging With MUSIC

The basic idea of the MUSIC algorithm is to perform an eigen-decomposition for the covariance matrix of CIR, resulting in a signal subspace orthogonal to a noise subspace corresponding to the signals reflected off the target. MUSIC is typically used for reconstructing the spatial spectrum of *sparse* signals. The reason why it is also applicable for imaging is that for each propagation delay $\tau_l$, the signals reflected off a target are sparsely distributed in the space. More specifically, as illustrated in Fig. 3, although the number of the reflected signals is large, these reflections occur over a large span of the propagation delays (i.e., ranges) and thus the number of signals with a certain propagation delay (i.e., reflected at a certain range) is small. Typically, there are only four to six significant reflected signals for each $\tau_l$. Therefore, for each $\tau_l$, the signal structure for target imaging is in line with the assumptions of the MUSIC algorithm, making the MUSIC algorithm feasible for solving the imaging problem. This is a result of utilizing the large bandwidth of the 60-GHz WiFi, which offers fine-grained range resolution.

Define a vector $\mathbf{h}_n(\tau_l) = [h^t_{1,n}(\tau_l), \ldots, h^t_{M,n}(\tau_l)]^T$, given a fixed transmit antenna $n$, to record the calibrated complex channel gains of all the receive antennas at a propagation delay $\tau_l$. To simplify the notations, we omit the dependence on the propagation delay $\tau_l$ and the transmit antenna index $n$ if not mentioned. Then, assuming that there are $D$ reflected signals impinging on the receive antenna array at the propagation delay $\tau_l$, the CIR $\mathbf{h}$ can be formulated as

$$\mathbf{h} = \left[\mathbf{s}(\theta_1, \phi_1), \ldots, \mathbf{s}(\theta_D, \phi_D)\right] \begin{bmatrix} x_1 \\ \vdots \\ x_D \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_M \end{bmatrix} \quad (2)$$

[4] Antennas are missing at four locations that are preserved for other purposes like power port, as shown in Fig. 2.

[5] This is because all the transmit/receive antennas share a single RF chain.

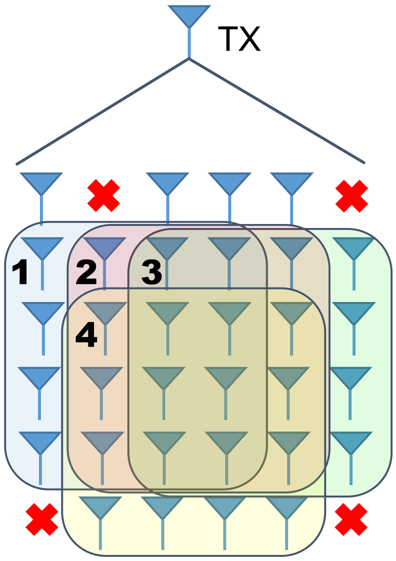



Fig. 4. SS. Four $4 \times 4$ subarrays are established from the $6 \times 6$ receive array.

where $\mathbf{s}(\theta_i, \phi_i)$ denotes the steering vector pointing to the direction $(\theta_i, \phi_i)$ corresponding to the incoming direction of the $i$th reflected signal, $x_i$ denotes the complex amplitude of that signal and $\varepsilon_i$ stands for additive noise, which is assumed to be Gaussian random variable with zero mean and independent and identically distributed (I.I.D.) for different receive antennas. More specifically, the steering vector $\mathbf{s}(\theta, \phi)$ records the phase response of the antenna array for a signal coming from the direction $(\theta, \phi)$ with its power normalized to 1, which can be expressed as

$$\mathbf{s}(\theta, \phi) = \frac{1}{\sqrt{M}} \begin{bmatrix} \Psi_\theta^0 \Omega_{\theta,\phi}^0 \\ \vdots \\ \Psi_\theta^{p-1} \Omega_{\theta,\phi}^{q-1} \\ \vdots \\ \Psi_\theta^{P-1} \Omega_{\theta,\phi}^{Q-1} \end{bmatrix} \tag{3}$$

where $\Psi_\theta$ and $\Omega_{\theta,\phi}$ are two basis functions defined as $\Psi_\theta \triangleq \exp(jkd \sin\theta)$ and $\Omega_{\theta,\phi} \triangleq \exp(jkd \cos\theta \sin\phi)$, $p$ and $q$ denote the row and column index of the antenna element on the array as shown in Fig. 2, $k$ is the wave number, and $d$ is the distance between two adjacent antennas along $y$ or $z$-axis. The indices without an antenna element (marked as red crosses in Fig. 4) are skipped in the steering vectors. A more concise matrix representation of (2) is written accordingly as

$$\mathbf{h} = S\mathbf{x} + \boldsymbol{\varepsilon} \tag{4}$$

where $S$ is defined as the *steering matrix*. Note that for a static target, the complex amplitude vector $\mathbf{x}$ is deterministic (fully coherent sources), and thus the covariance matrix of $\mathbf{h}$ would only contain the information of the noise. Therefore, the correlation matrix is used instead,[6] which can be expressed accordingly as

$$\begin{aligned} R &= \mathbb{E}[\mathbf{h}\mathbf{h}^H] \\ &= S\mathbf{x}\mathbf{x}^H S^H + \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^H] \\ &\triangleq R_s + R_\varepsilon \end{aligned} \tag{5}$$

where $R_s$ and $R_\varepsilon$ denote the correlation matrix for the signal components and noise, respectively. The eigenvalues $\lambda_1, \ldots, \lambda_M$ of $R$ are sorted in a nondescending order, associated with $M$ eigenvectors $\mathbf{e}_1, \ldots, \mathbf{e}_M$. Then, the noise subspace can be constructed as $E_\varepsilon = [\mathbf{e}_{D+1}, \ldots, \mathbf{e}_M]$, where $D$ stands for the rank of $R_s$, or, namely, the dimension of the signal subspace. The (pseudo) spatial spectrum for any direction $(\theta, \phi)$ can be obtained as

$$P(\theta, \phi) = \frac{1}{\mathbf{s}^H(\theta, \phi) E_\varepsilon E_\varepsilon^H \mathbf{s}(\theta, \phi)}. \tag{6}$$

Large values of the spatial spectrum $P(\theta, \phi)$ in a specific part of the space would most likely indicate the presence of one or more reflected signals; low values of $P(\theta, \phi)$ would indicate the absence of such reflections.

*Remark 1:* Note that the CBF-based spatial spectrum for any direction $(\theta, \phi)$ can be obtained as

$$P_{\mathrm{CBF}}(\theta, \phi) = \mathbf{s}^H(\theta, \phi) R \mathbf{s}(\theta, \phi) \tag{7}$$

and the MVDR-based spatial spectrum can be expressed as

$$P_{\mathrm{MVDR}}(\theta, \phi) = \frac{1}{\mathbf{s}^H(\theta, \phi) R^{-1} \mathbf{s}(\theta, \phi)}. \tag{8}$$

*Order Selection:* Another critical problem to apply MUSIC is to determine the number of signals $D$ that impinge on the array. In mmEye, Akaike information criterion (AIC) [29], a well-known information-theoretic approach for model order selection, is used. AIC is composed of two terms: 1) a data term, measuring the likelihood of the data given a certain $D$ and 2) a penalty term, measuring the complexity of the model. Specifically, $D$ is calculated as

$$\begin{aligned} D^\star = \arg\max_D \log &\left( \frac{\prod_{i=D+1}^{M} \lambda_i^{1/(M-D)}}{1/(M-D) \sum_{j=D+1}^{M} \lambda_j} \right)^{(M-D)} \\ &- D(2M - D) \end{aligned} \tag{9}$$

where $\lambda_i$ denotes the $i$th largest eigenvalue of the correlation matrix $R$. Since the AIC criterion tends to overestimate the number of impinging signals, AIC can retain the weak reflected signals to the greatest extent possible, which is desirable in the imaging application.

The MUSIC algorithm requires the rank of $R_s$ to be the same as the number of incoming signals $D$. However, since the rank of $R_s$ is only 1 which is likely much smaller than $D$, the performance of the MUSIC algorithm would deteriorate greatly or even completely fail to produce an effective spatial spectrum. To solve the problem, SS [22], [30], a commonly used technique for the rank deficiency issue, is applied as follows.

[6]Note that in this work, $\mathbf{h}$ is treated as a random vector and each experiment is just one realization of it. Under the assumption that the ensemble mean of $\mathbf{h}$ is equal to zero, i.e., $\mathbb{E}[\mathbf{h}] = \mathbf{0}$, the correlation matrix is equivalent to the covariance matrix.

### *B. Spatial Smoothing*

The idea of the SS is to split the receive array into several overlapping subarrays that share the same steering vectors except for certain angular rotations due to the differences in the ToA of the reflected signals impinging on different subarrays. Fig. 4 shows an example of the selected subarrays from the original $6 \times 6$ receive antenna array. As seen, due to the issue of missing antennas at certain locations of the array, no subarray with dimension $5 \times 5$ can be found and only four $4 \times 4$ antenna subarrays can be established.[7] Let $\mathbf{s}_{[k]}(\theta, \phi)$ denote the steering vector for the $k$th subarray, then we have $\mathbf{s}_{[2]}(\theta, \phi) = \Omega_{\theta,\phi}\mathbf{s}_{[1]}(\theta, \phi)$, $\mathbf{s}_{[3]}(\theta, \phi) = \Omega^2_{\theta,\phi}\mathbf{s}_{[1]}(\theta, \phi)$, and $\mathbf{s}_{[4]}(\theta, \phi) = \Psi_\theta \Omega_{\theta,\phi}\mathbf{s}_{[1]}(\theta, \phi)$. The correlation matrix of each subarray can be averaged to form the "spatially smoothed" correlation matrix $\tilde{R}$ with a higher rank, i.e.,

$$\tilde{R} = \frac{1}{K}\sum_{k=1}^{K} R_{[k]} \tag{10}$$

where $R_{[k]}$ denotes the correlation matrix of the $k$th subarray. It is proved in [23] that the rank of $\tilde{R}$ increases by 1 with probability 1 for each additional subarray in the averaging until it reaches its maximum value. Therefore, the rank of $\tilde{R}$ can be restored to 4 after the SS, which, however, is still under rank deficiency. To further solve the rank deficiency issue and reduce the variance of the spatial spectrum estimation, an exponential smoothing filter, which utilizes the time diversity of consecutive measurements, is applied to the estimation of the correlation matrix

$$\tilde{R}_t = \beta\tilde{R}_{t-1} + (1-\beta)\tilde{R} \tag{11}$$

where $\beta$ is the smoothing factor. The value of $\beta$ is chosen based on the tradeoff between the responsiveness and accuracy of the system, and mmEye uses $\beta = 0.9$ for the SS-based method. The spatial spectrum for each $\tau_l$ can be thus produced by (6).

The idea of SS is also implemented in ArrayTrack [31] and SpotFi [32] for active target localization with WiFi. However, they perform SS along with 1-D array and subcarriers and aim to detect only the AoA of the direct path signal. Differently, mmEye performs 2-D SS and targeting at imaging, which needs to identify all reflection signals.

### *C. Joint Transmitter Smoothing*

Although SS can improve the performance of the MUSIC algorithm under highly correlated sources, it reduces the effective aperture of the array (changed from $6 \times 6$ to $4 \times 4$), which equivalently increases the beamwidth of the array and decreases the spatial resolution. Could we solve the rank deficiency problem of the correlation matrix without the loss of the antenna aperture at the same time? We offer an affirmative answer in mmEye by exploiting the Tx diversity and accordingly devising a novel JTS technique.

As the AoA of each reflected signal is only relative to the receive array, each receive array corresponding to each Tx antenna should share the same set of the steering vectors except for the angular rotations, similar to the discussions in the above section for classical SS. However, the angular rotation is not due to the shifting of the subarrays at the Rx array, but instead is caused by the tiny differences in the locations of the Tx antennas. Considering the small wavelength, these tiny differences can generate significant enough phase deviations to the signals received by the receive array coming from different TX antenna, which enables SS across the receive arrays associated with different TX antennas. Thus, it is feasible to treat the whole receive antenna array, for each specific TX antenna, as a subarray. Thus, in total, $N$ subarrays can be formulated.

Recall that $\mathbf{h}_n(\tau_l)$ denotes the received CIR at $\tau_l$ for the $n$th transmit antenna. Define the channel matrix for each $\tau_l$ as $H(\tau_l) = [\mathbf{h}_1(\tau_l), \ldots, \mathbf{h}_N(\tau_l)]$. Then, the corresponding correlation matrix at $\tau_l$ after SS can be obtained as

$$\begin{aligned}\tilde{R}_{Tx}(\tau_l) &= \frac{1}{N}\sum_{n=1}^{N}\mathbf{h}_n(\tau_l)\mathbf{h}_n^H(\tau_l)\\ &= \frac{1}{N}H(\tau_l)H^H(\tau_l).\end{aligned} \tag{12}$$

$\tilde{R}_{Tx}(\tau_l)$ is now a full-rank matrix and multiple measurements are not required unlike the case for SS based on a single transmit antenna, which increases the responsiveness of mmEye greatly. Nevertheless, the exponential filter can still improve the robustness of the spatial spectrum estimation, which can be produced by (6). mmEye uses $\beta = 0.5$ for the JTS method, which implies that mmEye needs 2 complete CIR recordings in average to construct an image. Considering the 15 Hz channel sounding rate of the device, it only takes mmEye about 0.13 s to capture an image, which is sufficient for realtime applications. Interestingly, the matrix $H(\tau_l)H^H(\tau_l)$ is also known as the *time-reversal matrix* [33]. If the Tx and Rx could share the same array, the proposed imaging algorithm is related to the time-reversal MUSIC (TR-MUSIC) imaging algorithm [34] with minor modifications to the formation of the steering vectors.

### *D. Joint Receiver Smoothing*

Thanks to the channel reciprocity [35], the imaging can also be performed at the transmitter side as well. By simply transposing the channel matrix $H(\tau_l)$, we can obtain another set of channel measurements $H^T(\tau_l)$ between the Tx and Rx antennas if the receive antennas were transmitting and the transmit antennas were receiving. Similarly, the corresponding correlation matrix after the joint receiver smoothing (JRS) at $\tau_l$ is obtained as $\tilde{R}_{Rx}(\tau_l) = 1/NH^T(\tau_l)H^*(\tau_l)$, where $(\cdot)^*$ denotes the conjugate operation. However, the quality of imaging on the Tx side is a little worse than that on the Rx side. This is because, during the channel sounding, the device first uses a fixed Tx antenna and scans through all the Rx antennas before switching to the next Tx antenna, which makes the phase measurements of different Tx antennas less coherent. Therefore, mmEye only utilizes the JTS technique in practice.

[7] A square subarray is just one example of the subarray, which has the merit that the spatial resolution for both azimuth and elevation are the same.

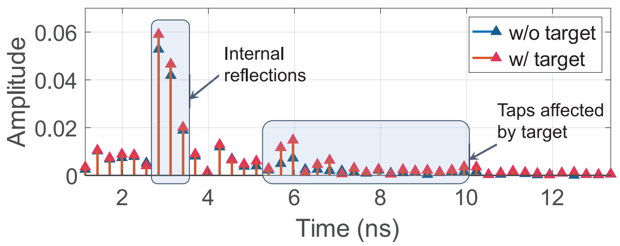



Fig. 5. Two illustrative CIRs: measured w/ and w/o a target in front of the device.

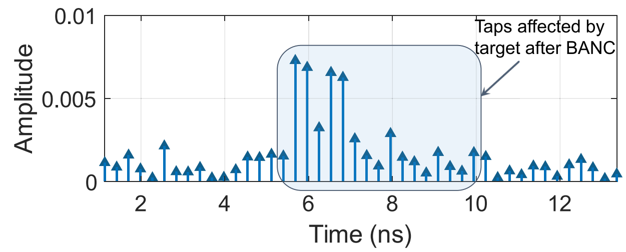



Fig. 6. Obtained CIR after BANC.

## IV. System Design

The work flow of mmEye is simple: put the device at a fixed location and perform a background calibration by collecting seconds of measurements, then the system is ready to image humans and objects present in the field of view. In this section, we present the design of a functional system based on the proposed super-resolution imaging algorithm. We mainly incorporate two additional components of BANC and target detection before the ultimate imaging.

### A. Background and Noise Cancelation

Besides the target of interest, the transmitted signals may also be reflected by the background objects, e.g., furniture, ceiling, grounds, walls, etc. In addition, there are internal signal reflections on the intermediate frequency (IF) cable connectors, as shown in Fig. 5. These undesired reflections from the background together with the internal noise interfere with the signals reflected off the target and thus degrade the imaging the quality. To combat these problems, in the following, a BANC algorithm is proposed to filter out the background reflections and the internal noise.

Mathematically, the CIR $h_{m,n}$ can be modeled as the sum of the target-related component $h^t_{m,n}$ and the background/internal reflection-related component $h^b_{m,n}$. To obtain $h^t_{m,n}$, mmEye first collects a bunch of CIRs for the background without the presence of the target to estimate $h^b_{m,n}$, and then obtains $h^t_{m,n}$ by subtracting $h^b_{m,n}$ from the newly measured CIR $h_{m,n}$ with the presence of the target.

Assume that there are $Q$ samples of the CIRs measured without the target. Then, $h^b_{m,n}$ can thus be estimated by the sample mean of the measured background CIRs, i.e., $h^b_{m,n}(\tau_l) \approx 1/Q\sum_{q=1}^{Q} h_{m,n}(\tau_l, t_q)$. Due to the synchronization errors and automatic gain control (AGC) module on the chip, the random amplitude and common initial phase of the CIRs changes from frame to frame. Therefore, it is not feasible to subtract the background CIR directly from the CIR with the target. A complex scaling factor $\alpha$ is thus applied to scale the obtained background CIR before the cancelation. The clean CIR $h^t_{m,n}$ after the BANC can be obtained accordingly as

$$h^t_{m,n}(\tau_l, t) = h_{m,n}(\tau_l, t) - \alpha h^b_{m,n}(\tau_l). \tag{13}$$

Regarding to the choice of $\alpha$, a minimum mean square error (MMSE) estimator is applied which selects the value of $\alpha$ that minimizes the energy of $h^t_{m,n}(\tau_l, t)$ over the first $L_0$ CIR taps, i.e.,

$$\alpha^\star = \arg\min_\alpha \sum_{l=0}^{L_0-1} |h_{m,n}(\tau_l, t) - \alpha h^b_{m,n}(\tau_l)|^2. \tag{14}$$

The analytical form of optimal solution $\alpha^\star$ can be derived correspondingly as

$$\alpha^\star = \frac{\sum_{l=0}^{L_0-1} [h^b_{m,n}(\tau_l)]^H h_{m,n}(\tau_l, t)}{\sum_{l=0}^{L_0-1} [h^b_{m,n}(\tau_l)]^H h^b_{m,n}(\tau_l, t)} \tag{15}$$

where $x^H$ denotes the Hermitian of $x$. The intuition for only using the first $L_0$ taps to estimate $\alpha$ is that the target being imaged is usually at a certain distance from the device to be observed completely in the specific field of view. Therefore, the first few taps are not affected by the presence of the target and are only affected by the AGC, leading to a more accurate estimation of the scaling factor. Fig. 6 shows an example of the CIR after the BANC. It can be observed that the impact of the target on the CIR taps has been greatly magnified in terms of amplitude.

*Remark 2:* If the random initial phase offset of the CIR is not compensated by BANC, the expectation of $R_s$ in (5) would be a zero matrix and the correlation matrix $R$ would only contain the noise component.

*Remark 3:* The background reflection signals could be affected when a target presents. However, this impact could be negligible. The affected background reflection signals usually travel a longer distance compared with those reflected off the target, Because the energy of the 60-GHz signals decay very fast as the propagation distance increases [36], the strength of the signals reflected from the target is much stronger than that of the affected background reflection signals. Besides, the target detection algorithm only selects the most dominant reflection signal for each direction. Therefore, the background subtraction algorithm works desirably for most of the cases.

### B. Target Detection

The purpose of object detection is to robustly detect all the CIR taps that are affected by the target(s) of interest. Because not all the RF signals reflected off human body parts can be captured by the receive antenna array, the energy of the signals reflected off some parts of the target can be very weak. To increase the "visibility" of the weak reflections, for each propagation delay, we calculate the variation of the energy distribution of the spatial spectrum $V_t(\tau)$, defined as $V_t(\tau) = \mathrm{Var}_\theta[\mathrm{Var}_\phi[P_t(\theta, \phi, \tau)]]$, where $\mathrm{Var}_\theta[\cdot]$ denotes

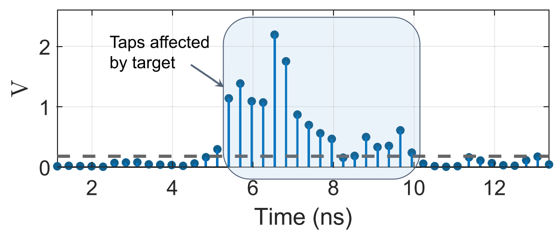



Fig. 7. RoI detection.

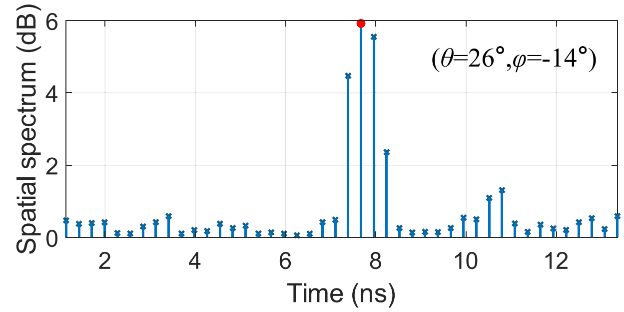
the variance over parameter $\theta$, $P_t(\theta, \phi, \tau)$ denotes the spatial spectrum for the direction $(\theta, \phi)$ and the propagation delay $\tau$ measured at time slot $t$. A large $V_t(\tau)$ implies that the energy distribution of the reflected signals for that range is highly nonuniform in space, indicating the presence of a target in that specific range, while for the range where no target presents, the energy of the reflected signals is usually small and uniformly distributed in space. By considering the variance over all the directions, the signals that may be very small along some specific directions could be amplified. Then, the set of the Range of Interest (RoI) at time slot $t$ are formed as $\text{RoI}(t) = \{\tau | V_t(\tau) > \eta, \ \forall \tau\}$, where $\eta(t)$ is a preset threshold. To accommodate the time-varying interference and noise, as illustrated in Fig. 7, we use a multiple of the median value of $V_t(\tau)$ as the threshold for each time slot $t$, i.e., $\eta(t) = \kappa \text{Med}_\tau[V_t(\tau)]$, where $\kappa$ denotes a constant coefficient, and $\text{Med}_\tau[\,\cdot\,]$ denotes the median value over $\tau$. The reason we use the median to determine the threshold is that the median of $V_t(\tau)$ can adaptively capture the variations of the noise level of the board especially when the total number of the taps $L$ is large.

Then, we only need to search for the Points of Interest (PoI) over the spatial spectrum $P(\theta, \phi, \tau)$ within the RoI set. Due to the fact that the millimeter wave cannot penetrate the general object well, e.g., millimeter waves are mostly absorbed within the human skin [37], only the first significant point of $P(\theta, \phi, \tau)$ w.r.t. $\tau$ contains the information of the target. Specifically, mmEye locates the PoI based on the following rule: given the spatial spectrum for each direction $(\theta, \phi)$, try to find the first local maximum point of $P(\theta, \phi, \tau)$ along $\tau$ within the RoI set that exceeds a preset threshold $\gamma$ (chose as a small value to filter out device noise); if failed, then no point of interest is found for this direction. For each point of interest $(\theta_i^\star, \phi_i^\star, \tau_i^\star)$, the associated weight is the value of the corresponding spatial spectrum, i.e., $P(\theta^\star, \phi^\star, \tau^\star)$.

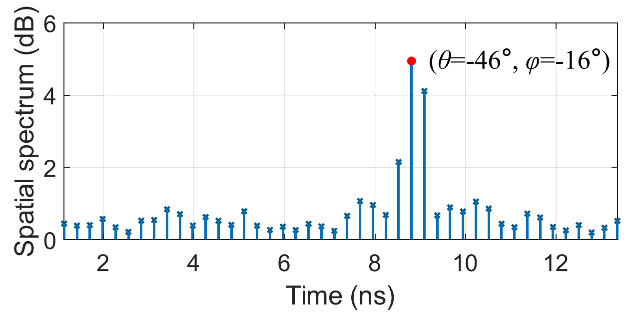
Fig. 8 shows two examples of the obtained spatial spectrum for different spatial directions. The red dot in both examples indicates the point of interest with weights 5.92 and 4.94 dB, respectively. The set of the PoI at time slot $t$ is denoted as $\text{PoI} = \{(\theta_i^\star, \phi_i^\star, \tau_i^\star, P_i^\star), i = 1, \ldots, S\}$, where $S$ is the total number of PoI and $P_i^\star$ denotes the value of the spatial spectrum corresponding to that point, i.e., $P_i^\star \triangleq P(\theta_i^\star, \phi_i^\star, \tau_i^\star)$.

### C. Imaging

Finally, we transform the PoI into plain images with depth and weighting information. mmEye first converts the PoI from the polar coordinates $(\theta_i^\star, \phi_i^\star, \tau_i^\star)$ to Cartesian coordinates $(x_i^\star, y_i^\star, z_i^\star)$ by applying simple geometric transformations.




(a)




(b)

Fig. 8. Examples of the spatial spectrum over $\tau$ for different directions. (a) Example #1. (b) Example #2.

Then, all the PoI are projected to a 2-D-plane that is parallel to the $y-z$ plane, as shown in Fig. 2, with a certain depth $x_d$, which is defined as the distance between these two planes. $x_d$ is optimized as below to ensure that the shape of the target in the obtained image agrees with the target's actual shape.

The optimal depth $x_d$ is determined automatically by solving a weighted least absolute deviation problem

$$x_d^\star = \arg\min_{x_d} \sum_{i=1}^{S} (P_i^\star - \gamma)(x_d - x_i^\star)^2 \tag{16}$$

which minimizes the $\ell_2$-norm of the distances between the PoI and the selected plane, weighted by their importance $(P_i^\star - \gamma)$, where $\gamma$ is the same threshold used in the target detection and thus the weights are always positive. It is designed to preserve the most of information of the PoI round the projected plane. To further remove the outliers within the set of PoI, mmEye only selects the points that are close enough to the projected plane, i.e., $|x_i^\star - x_d^\star| \le w$, where $w$ is a preset threshold.

Fig. 9 portrays two examples of the obtained image of a person, which shows that the proposed super-resolution algorithm significantly outperforms prior approaches CBF and MVDR and achieves comparable results with Kinect.

## V. Evaluation

In this section, we evaluate mmEye in practical settings using a commodity 802.11ad chipset. We study the imaging quality for both humans and objects and both LOS and NLOS. We also compare mmEye with existing beamforming techniques CBF and MVDR.

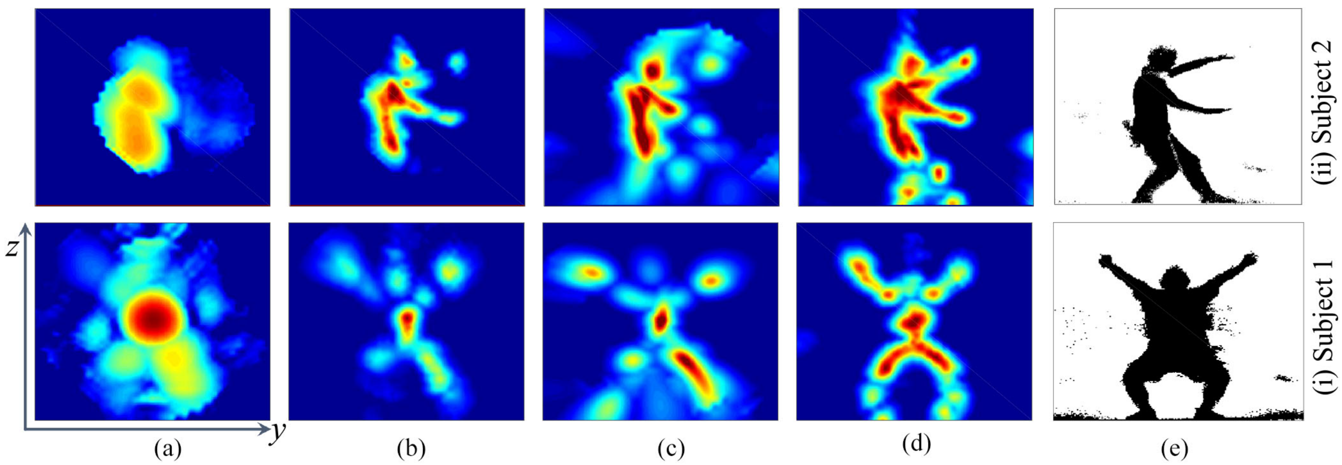



Fig. 9. Examples of the imaging results. (a)–(d) show the examples of the obtained images by different spatial spectrum estimators. The color indicates the value of $P_i^{\star}$ for the $i$th detected point, and the higher value, the redder color (c) and (d) are the results by mmEye with SS and JTS, respectively (e) is obtained as the ground truth by a Kinect depth sensor.

## A. Methodology

*Experiment Setup:* We prototype mmEye and conduct real-world experiments using a Qualcomm 802.11ad chipset. The chipset is equipped with two antenna arrays, both having 32 antennas arranged in a 6×6 topology. During experiments, the device is operating in a radar mode, i.e., the Tx antennas constantly transmit pulses and the Rx antennas receive the reflected signals and estimate the CIR accordingly. The channel sounding rate of the device is set to 15 Hz.

Our experiments take place on one floor of a typical office building of size 28 m ×36 m, which is furnished with desks, chairs, computers, and TVs. A typical setup of the system is shown in Fig. 2. Both humans and everyday objects are tested in our experiment. For human imaging, we recruit 4 volunteers (two males and two females) and test out at different locations and distances with different postures. We mainly focus on quasi-static scenarios but also test for moving targets. Our evaluation consists of both single-person case and multiple person case. For static cases, each subject performs 10 to 15 different postures as he/she will and we collect about 30 s of data for each posture. mmEye runs in realtime and outputs an image for every four CIR recordings. For object imaging, we test with everyday objects, such as fans, heaters, monitors, suitcases, etc., that have various shapes, sizes, and materials.

*Ground Truth:* Prior works that focus on imaging static objects mainly assume known ground truths of the objects [10], [16], [38], [39]. To evaluate human imaging, however, we could not obtain ground truth from manual measurements of the target dimensions and shapes. Instead, we extract images from a Kinect depth sensor by the library provided by [40] to serve as ground truth. To detect the target of interest from a Kinect frame, we simply search for a certain depth and extract all points in that depth. One could perform advanced segmentation by combining the RGB sensor for this purpose, which is however out of the scope of this article. Note that the measurements and target detection on Kinect both contain noises, which do not favor our evaluation. The results by mmEye and Kinect are shifted and interpolated so that their coordinates are aligned with identical point density.

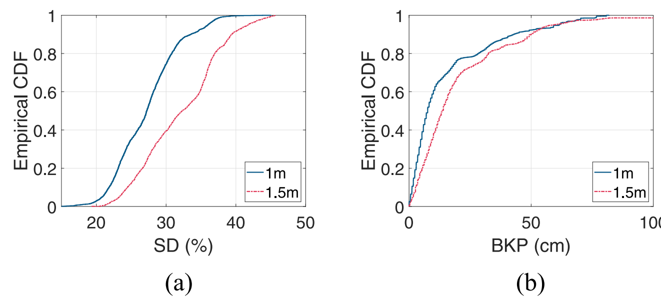



Fig. 10. Performance of human imaging over subject-to-device distance. (a) SD. (b) BKP.

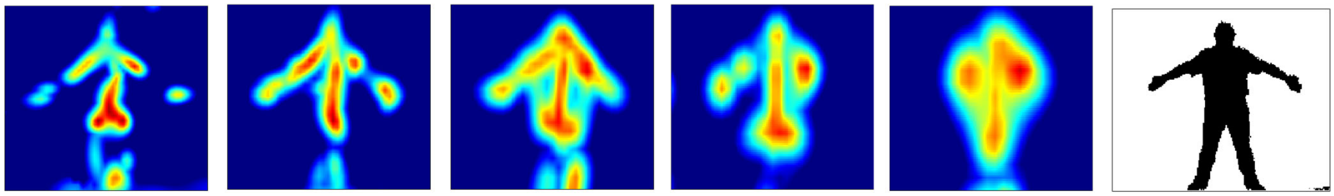
*Evaluation Metrics:* It is not easy to define proper metrics to evaluate the imaging quality, although it is intuitive for a human to tell from the visual results. In addition to qualitative visual comparisons, we propose two quantitative metrics in this work as follows.

1) *Silhouette Difference (SD):* The percentage of XOR difference between the mmEye images (after thresholding) and the Kinect frames, ranging from 0 (no errors) to 1 (completely different).
2) *Boundary Key-Point Precision (BKP):* The absolute location error for several key points on the target boundary. Since we do not have labeled points for Kinect and mmEye, we mainly account for the topmost, leftmost, and rightmost points in our evaluation, which can be automatically detected.

## B. Performance

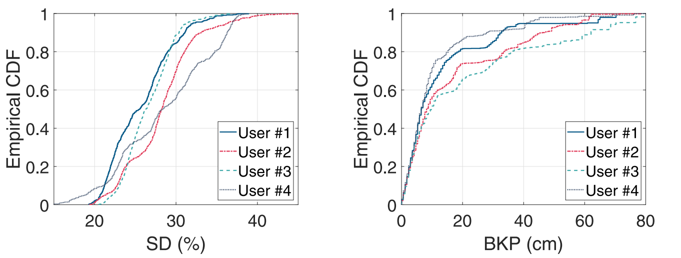
We now evaluate the imaging quality. Since we cannot include all the visual imaging results in this article, the images and videos of our evaluation are provided in [41] for online access.

*1) Human Imaging:* We first evaluate imaging performance for human targets. Figs. 1 and 9 already illustrate some of the visual results, which evidently demonstrate the remarkable



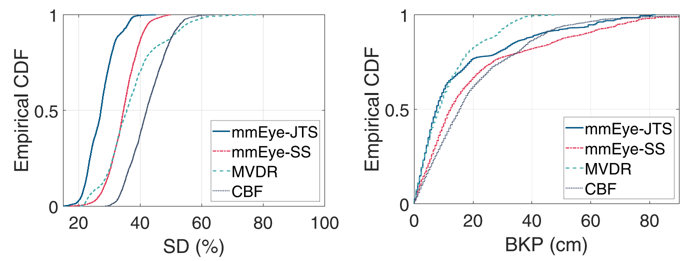
(a) (b) (c) (d) (e) (f)

Fig. 11. Impact of distance on human imaging. A subject stands in front of the device with varying distances and performs the same posture, as illustrated in the Kinect depth image. (a) 1m. (b) 1.5m. (c) 2m. (d) 2.5m. (e) 3m. (f) Kinect.




(a) (b)

Fig. 12. Performance comparison across different users with respect to (a) SD and (b) BKP.




(a) (b)

Fig. 13. Performance comparison among different spatial spectrum estimators. (a) SD. (b) BKP.

imaging quality achieved by mmEye. Now we quantitatively evaluate the precision over all the testing data using the SD and BKP metrics.

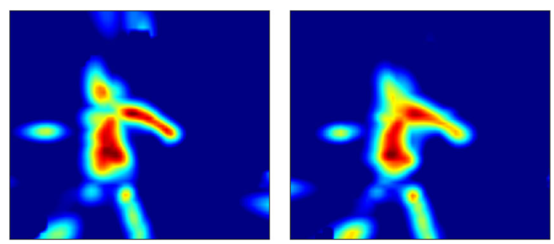
As shown in Fig. 10(a) and (b), mmEye achieves the median of 27.2% for SD and 7.6 cm for BKP when subjects are about 1m away from the device; while it degrades to the median of 32.3% for SD and 13.5 cm for BKP when subjects are about 1.5 m away. This is mainly because a larger distance between the target and device leads to a wider beam and a weaker reflected signal, both affecting imaging quality.

To visualize the degradation of imaging resolution for different distances between the subject and the device, we let a subject, performing the same posture, stand in front of the device with varying distances ranging from 1 to 3 m. The corresponding imaging result, as illustrated in Fig. 11, shows that as the distance increases, the resolution of imaging degrades gradually and the contour of the human body becomes blurry especially when the distance is larger than 2 m.

*User Diversity:* Fig. 12(a) and (b) show the imaging quality of mmEye for different persons w.r.t. SD and BKP, respectively. The results show consistently accurate imaging for different subjects. The slight variations in performance are due to that the body type and clothing are varying among the subjects, which can affect the strength of the RF signals reflected off the human body.

*Performance Comparison:* Now we show the super-resolution performance of mmEye by comparing it with existing beamforming techniques, including CBF and MVDR. We also implement and compare two variations of mmEye, i.e., mmEye with SS (SS, as in Section III-B) and mmEye with JTS (JTS, as in Section III-C) to show the considerable benefits of the proposed JTS algorithm.

Fig. 9 shows the visual results of two subjects using different methods. As seen, mmEye-JTS achieves the best imaging results comparable with Kinect, while mmEye-SS stands the second-best yet is already much worse than mmEye-JTS. MVDR can see parts of the human body but misses many others, while CBF does not capture body parts but only detects the human body as a whole.



(a) (b)

Fig. 14. Imaging on (a) Tx/ (b) Rx array. The quality of the image obtained on Rx array is better than that obtained on the Tx array.

Further in Fig. 13, we show the quantitative results of different approaches. As for the SD metric, mmEye-JTS achieves the best performance and mmEye-SS comes in the second place, which agrees with the visual results shown in Fig. 9. Note that MVDR performs better than other techniques w.r.t. BKP metric, however, it performs poorly when regarding to SD metric. This is because the spatial spectrum estimation of MVDR is more conservative and thus it misses some of the major parts of the human body, which does not necessarily increase errors in BKP (e.g., the topmost point does not change too much). In principle, only good results in both metrics indicate good quality of imaging.

In addition, the comparison between the JTS and JRS, as discussed in Sections III-C and III-D, is shown in Fig. 14. It shows that the quality of the image obtained on Rx array is better than that obtained on the Tx array due to the reason discussed in Section III-D.

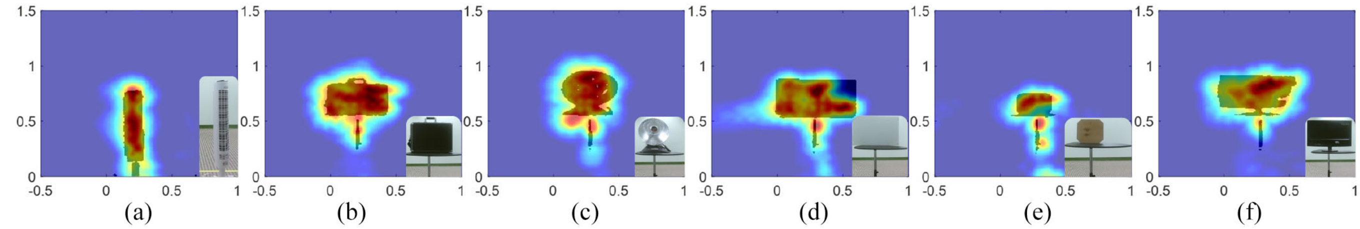


Fig. 15. Imaging for different static objects. Objects with different shapes, sizes, materials, and surface conditions are tested. The objects are placed about 1.5 m away from the device. The shadow overlays show the ground truths obtained by Kinect. (a) Fan. (b) Plastic suitcase. (c) Heater. (d) Large wood panel. (e) Small wood panel. (f) Monitor.

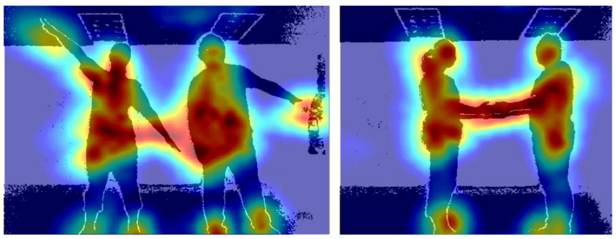


Fig. 16. Imaging for multiple persons. The two persons stand about 1.5 m away from the device. The ground truths are displayed in the overlay.

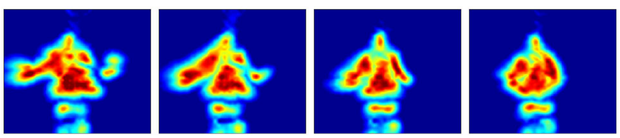


Fig. 17. Imaging for a person in motion. A standing person puts down his arms at normal speed. Videos for continuous imaging can be found in [41].

*2) Object Imaging:* mmEye can also image objects. We test real objects of different shape, size, curvature, surface, and material. Fig. 15 shows some of the testing objects with mmEye imaging results, Kinect ground truths, and pictures displayed. The objects we select reveal different shapes (cylinder, square, and circle), materials (plastic, wood, and metal), size (from about 20 to 100 cm in length) As seen, mmEye can accurately image various objects. Specifically, mmEye achieves a median accuracy of 8.0 cm in shape estimation of the objects. The results show that mmEye achieves consistent performance for both human targets and objects.

*3) Case Studies:* In the following, we show how mmEye performs under different scenarios.

*Multiperson Imaging:* Actually, we already involve multiple objects in object imaging (Section V-B2) since we put the targets on a standing table. Yet we are more interested in multiple person imaging. Fig. 16 shows two imaging examples of two subjects with Kinect overlay as ground truths. As seen, both human figures are well captured, with the heads, feet, arms, and hands confidently recognized. The results underpin various applications of mmEye like multiuser gaming and user activity analysis.

*Dynamic Person Imaging:* Thanks to the JTS, mmEye can achieve imaging with one single snapshot but does not need successive measurements. Thus, it can effectively image targets in motion. We test both walking and in-place motion and show some visual imaging results in Fig. 17 where a user is moving arms. Not only does mmEye image the stationary body parts (e.g., the torso, and legs), it also tracks the moving parts (e.g., arms) successfully.

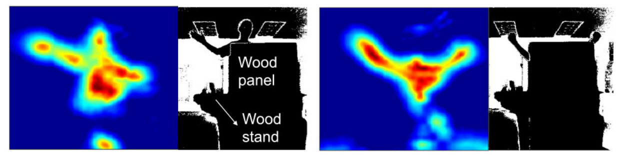



Fig. 18. Non-line-of-sight imaging. Left: The standing subject slightly raises his right hand (as seen from the figure) and stretches the left forearm horizontally; Right: The subject is in a similar posture as in the bottom row of Fig. 9, yet behind a big wood panel.

*Through-the-Wall Imaging:* While 60-GHz signals typically do not penetrate most obstacles, it is of great interest to examine if mmEye can image a target behind a thin drywall. We set up a large wood panel supported by a wood stand to emulate a drywall and test mmEye's performance under this setting. To better validate the performance, we ask the subject to expose partial of the body to the devices (mmEye and Kinect). As shown in Fig. 18, surprisingly, mmEye still captures a human figure behind the "drywall," while the Kinect depth sensor, as a vision-based sensor, is completely occluded by the big wood panel and only sees the exposed parts (i.e., hands in both images). The reason that mmEye can see through the wood panel is that the 60-GHz signals can penetrate the panel and reflect off the human body behind it. Albeit the reflected signals are much weaker, mmEye is still able to capture them by the effective BANC algorithm. However, we observe that the performance does degenerate in NLOS case. For example, the legs and feet in both figures are partially missing. It is more obvious when comparing the right figure to the bottom row in Fig. 9, where the subjects perform a similar posture. How to enhance the signals and overcome noises in through-the-wall scenarios remains an attractive problem.

*Human Face Imaging:* To show the high imaging resolution when the target is close, we ask a person to sit 30 cm away from mmEye and face the device. The imaging results are shown in Fig. 19, and the different head orientations of the head of the person can be clearly observed. This shows the potential of mmEye in applications that require high imaging resolution.

## VI. Discussion and Future Work

*Discussion:* mmEye takes an important step toward super-resolution imaging on 60-GHz WiFi radios. There are several limitations though. First, the working range of mmEye is

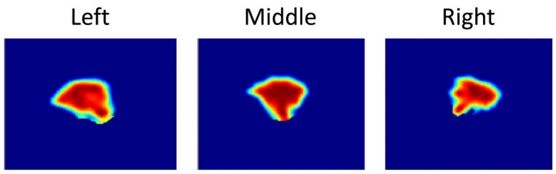



Fig. 19. Imaging for a human face. A person sits 30 cm away from the device and faces the device. Left: The person tilts the head to the left; Middle: The person keeps the head straight; Right: The person tilts the head to the right.

inherently determined by the 60-GHz radio. While our experimental device is specified to support up to 10 m for target detection, the imaging resolution decreases linearly over the range. Second, since 60-GHz signals can hardly penetrate walls or most objects, mmEye mainly images objects in a line-of-sight view but does not perform well for through-the-wall targets. How to extend the imaging range and enhance in NLOS scenarios is an immediate next Finally. Lastly, the imaging resolution of mmEye degrades with the distance between the object and the device increasing, resulting in a limited range for high-resolution imaging. When a subject is within 2 m, the skeleton structure of human body can been observed clearly from the RF images as shown in Fig. 11, and thus the applications requiring high imaging resolution, such as human pose estimation and human recognition, can be supported by mmEye. When a subject is far away from the device (> 2 m), the applications requiring moderate imaging resolution, such as multiperson tracking and people counting, can still be enabled.

*Future Work:* There are also multiple future directions. The combination of the JTS and JRS techniques could alleviate the specularity problem and improve the responsiveness of the system. Extension to SAR to increase the antenna aperture would further improve the resolution upon the super-resolution algorithm of mmEye. The performance for multiperson cases could also be further optimized. It is promising to study gesture recognition and multiuser gaming. Given the imaging results already achieved, it is of interest to study “wireless vision” problems like target segmentation, pose estimation, and human identification without face recognition, etc.

## VII. Related Work

*RF-Based Imaging:* RF imaging, particularly WiFi imaging, has been studied with great interest. WiFi signals have been exploited for various passive sensing, including motion tracking [6], [42], activity and gesture classification [3], [4], and material detection [9]. Object imaging (in outdoor space) with unmanned WiFi robots has been extensively studied [8], [43], which however, does not apply to indoor space. Indoor WiFi imaging using commodity WiFi signals is studied in [10], [38] with emulated large arrays. The resolution is limited by inherent WiFi signals on 2.4 GHz/5 GHz bands [10]. To improve the precision on WiFi bands, specialized FMCW radar is designed with large phased array to capture human figs. [7] and pose information [11], [12]. TagScan [39] employs RFID for static object imaging and material identification, but needs dedicated hardware (i.e., RFID reader) to be deployed on both sides of the target. 60-GHz radios are recently used for precise tracking [26], [44], vital sign monitoring [27], [45], material sensing [46], and object navigation [47]. A recent work [16] exploits RSS analysis to image objects using a pair of 60-GHz devices with the receiver moving. It only images objects but not humans. And it uses a horn antenna for steering control, leaving the rich phase information and phased array beamforming uncharted. Differently, our work exploits an unseen opportunity of radar processing, a dual role of emerging 60-GHz networking chipsets, and enables super-resolution imaging for both humans and objects by using the built-in phased arrays.

Recently, deep learning has been exploited for RF sensing. A gesture recognition model is proposed in [4] with a large-scale WiFi data set, which is publicly available and would play an important role to the community. Advances in RF imaging have achieved both 2-D and 3-D pose construction from WiFi signals via neural networks [48], [49]. However, these works require extensive training and multiple pairs of transceivers. In contrast, mmEye offers a training-free approach based on a single radio.

*Camera-Based Imaging:* Nowadays camera is increasingly popular for object recognition and segmentation [50]–[54]. While great success has been made for 2-D imaging, to obtain position and depth information, however, usually requires RGB-Depth cameras like Kinect sensors [25] and camera arrays like VICON [55]. Also, camera-based solutions depend on lighting conditions and are privacy sensitive. As a comparison, mmEye aims to enable a depth camera by reusing a commodity networking device, which works without any light and preserves privacy.

*Sonar and Radar Systems:* Radar systems have been studied for decades and can achieve high-precision imaging of objects and humans [13], [56]. These systems, however, usually use terahertz [57], laser [58], or millimeter and submillimeter waves [13], [59], [60] on special hardware with large aperture, which are not suitable for ubiquitous applications. Albeit mmEye also follows a radar-like operation to measure CIR, it differs by reusing commodity 60-GHz networking chipsets and contributes novel and different algorithms to overcome the unique challenge of an extremely small aperture.

## VIII. Conclusion

This article presents mmEye, a super-resolution imaging system toward a mmWave camera on commodity 60-GHz WiFi devices. mmEye contributes a novel super-resolution imaging algorithm based on MUSIC with JTS. Experiments show that mmEye achieves remarkable imaging performance for both humans and everyday objects, comparable to that of dedicated depth sensors like Kinect. mmEye can even image a person behind a thin drywall. We believe mmEye takes an important step toward ubiquitous imaging and inspires research on 60-GHz WiFi sensing.

## References

[1] B. Wang, Q. Xu, C. Chen, F. Zhang, and K. R. Liu, “The promise of radio analytics: A future paradigm of wireless positioning, tracking, and sensing,” *IEEE Signal Process. Mag.*, vol. 35, no. 3, pp. 59–80, May 2018.

[2] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Comput. Surveys*, vol. 52, pp. 1–36, Jun. 2019.
[3] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. ACM 21st Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2015, pp. 65–76.
[4] Y. Zheng *et al.*, "Zero-effort cross-domain gesture recognition with Wi-Fi," in *Proc. 17th Annu. Int. Conf. Mobile Syst. Appl. Serv.*, 2019, pp. 313–325.
[5] F. Zhang *et al.*, "SMARS: Sleep monitoring via ambient radio signals," *IEEE Trans. Mobile Comput.*, early access, Sep. 6, 2019, doi: 10.1109/TMC.2019.2939791.
[6] F. Adib, Z. Kabelac, and D. Katabi, "Multi-person localization via RF body reflections," in *Proc. 12th USENIX Conf. Netw. Syst. Design Implement. (NSDI)*, 2015, pp. 279–292.
[7] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand, "Capturing the human figure through a wall," *ACM Trans. Graph.*, vol. 34, no. 6, p. 219, 2015.
[8] C. R. Karanam and Y. Mostofi, "3D through-wall imaging with unmanned aerial vehicles using WiFi," in *Proc. 16th ACM/IEEE Int. Conf. Inf. Process. Sens. Netw. (IPSN)*, Pittsburgh, PA, USA, 2017, pp. 131–142.
[9] C. Wang, J. Liu, Y. Chen, H. Liu, and Y. Wang, "Towards in-baggage suspicious object detection using commodity WiFi," in *Proc. IEEE Conf. Commun. Netw. Security (CNS)*, Beijing, China, 2018, pp. 1–9.
[10] D. Huang, R. Nandakumar, and S. Gollakota, "Feasibility and limits of Wi-Fi imaging," in *Proc. 12th ACM Conf. Embedded Netw. Sens. Syst.*, 2014, pp. 266–279.
[11] M. Zhao *et al.*, "Through-wall human pose estimation using radio signals," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 7356–7365.
[12] M. Zhao *et al.*, "RF-based 3D skeletons," in *Proc. Conf. ACM Spec. Interest Group Data Commun. (SIGCOMM)*, 2018, pp. 267–281.
[13] R. Appleby and R. N. Anderton, "Millimeter-wave and submillimeter-wave imaging for security and surveillance," *Proc. IEEE*, vol. 95, no. 8, pp. 1683–1690, Aug. 2007.
[14] J. Gollub *et al.*, "Large metasurface aperture for millimeter wave computational imaging at the human-scale," *Sci. Rep.*, vol. 7, Feb. 2017, Art. no. 42650.
[15] V. M. Patel, J. N. Mait, D. W. Prather, and A. S. Hedden, "Computational millimeter wave imaging: Problems, progress, and prospects," *IEEE Signal Process. Mag.*, vol. 33, no. 5, pp. 109–118, Sep. 2016.
[16] Y. Zhu, Y. Zhu, B. Y. Zhao, and H. Zheng, "Reusing 60GHz radios for mobile radar imaging," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2015, pp. 103–116.
[17] (2019). *Qualcomm 802.11ad 60GHz WiFi*. [Online]. Available: https://www.qualcomm.com/solutions/networking/features/80211ad
[18] (2019). *Talon AD7200 Multi-Band Wi-Fi Router*. [Online]. Available: https://www.tp-link.com/us/home-networking/wifi-router/ad7200/
[19] (2018). *Asus Foreshadows 5G Handsets Featuring 802.11ad WiFi*. [Online]. Available: https://www.eetimes.com/author.asp?section_id=36&doc_id=1333373
[20] (2018). *60-GHz Wi-Fi Gets a Refresh: Qualcomm Rolls Out Chips for 802.11ay*. [Online]. Available: https://www.eetimes.com/document.asp?doc_id=1333870
[21] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
[22] H. Wang and K. R. Liu, "2-D spatial smoothing for multipath coherent signal separation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 2, pp. 391–405, Apr. 1998.
[23] J. H. Cozzens and M. J. Sousa, "Source enumeration in a correlated signal environment," *IEEE Trans. Signal Process.*, vol. 42, no. 2, pp. 304–317, Feb. 1994.
[24] M. F. Iskander, *Electromagnetic Fields and Waves*. Long Grove, IL, USA: Waveland Press, 2013.
[25] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, Apr. 2012.
[26] T. Wei and X. Zhang, "mTrack: High-precision passive tracking using millimeter wave radios," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2015, pp. 117–129.
[27] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, "Monitoring vital signs using millimeter wave," in *Proc. 17th ACM Int. Symp. Mobile Ad Hoc Netw. Comput. (MobiHoc)*, 2016, pp. 211–220.
[28] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.
[29] A. P. Liavas and P. A. Regalia, "On the behavior of information theoretic criteria for model order selection," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1689–1695, Aug. 2001.
[30] T.-J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 4, pp. 806–811, Aug. 1985.
[31] J. Xiong and K. Jamieson, "ArrayTrack: A fine-grained indoor location system," in *Proc. 10th USENIX Conf. Netw. Syst. Design Implement. (NSDI)*, 2013, pp. 71–84.
[32] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "SpotFi: Decimeter level localization using WiFi," in *Proc. ACM SIGCOMM Comput. Commun. Rev.*, vol. 45, pp. 269–282, Aug. 2015.
[33] C. Prada, S. Manneville, D. Spoliansky, and M. Fink, "Decomposition of the time reversal operator: Detection and selective focusing on two scatterers," *J. Acoust. Soc. Amer.*, vol. 99, no. 4, pp. 2067–2076, 1996.
[34] A. J. Devaney, "Time reversal imaging of obscured targets from multistatic data," *IEEE Trans. Antennas Propag.*, vol. 53, no. 5, pp. 1600–1610, May 2005.
[35] Z.-H. Wu, Y. Han, Y. Chen, and K. R. Liu, "A time-reversal paradigm for indoor positioning system," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1331–1339, Apr. 2015.
[36] K. Haneda, J. Järveläinen, A. Karttunen, M. Kyrö, and J. Putkonen, "A statistical spatio-temporal radio channel model for large indoor environments at 60 and 70 GHz," *IEEE Trans. Antennas Propag.*, vol. 63, no. 6, pp. 2694–2704, Jun. 2015.
[37] S. Alekseev, O. Gordiienko, and M. Ziskin, "Reflection and penetration depth of millimeter waves in murine skin," *Bioelectromagnetics*, vol. 29, no. 5, pp. 340–344, 2008.
[38] P. M. Holl and F. Reinhard, "Holography of Wi-Fi radiation," *Phys. Rev. Lett.*, vol. 118, no. 18, 2017, Art. no. 183901.
[39] J. Wang, J. Xiong, X. Chen, H. Jiang, R. K. Balan, and D. Fang, "TagScan: Simultaneous target imaging and material identification with commodity RFID devices," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2017, pp. 288–300.
[40] L. Xiang *et al.* (2016). *libfreenect2: Release 0.2*. [Online]. Available: http://github.com/OpenKinect/libfreenect2
[41] (2019). *mmEye Images and Videos*. [Online]. Available: https://www.dropbox.com/sh/7kyhywk92rnjtp1/AADLRd7zaNPAh7sDcOLhv1cMa?dl=0
[42] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu, "Widar2.0: Passive human tracking with a single Wi-Fi link," in *Proc. 16th Annu. Int. Conf. Mobile Syst. Appl. Serv. (MobiSys)*, 2018, pp. 350–361.
[43] S. Depatla, L. Buckland, and Y. Mostofi, "X-ray vision with only WiFi power measurements using Rytov wave models," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1376–1387, Apr. 2015.
[44] C. Wu, F. Zhang, B. Wang, and K. R. Liu, "mmTrack: Passive multi-person localization using commodity millimeter wave radio," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, Toronto, ON, Canada, 2020, pp. 2400–2409.
[45] F. Wang, F. Zhang, C. Wu, B. Wang, and K. R. Liu, "ViMo: Multi-person vital sign monitoring using commodity millimeter wave radio," *IEEE Internet Things J.*, early access, Jun. 22, 2020, doi: 10.1109/JIOT.2020.3004046.
[46] C. Wu, F. Zhang, B. Wang, and K. R. Liu, "mSense: Towards mobile material sensing with a single millimeter-wave radio," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 4, no. 3, pp. 1–20, 2020.
[47] Y. Zhu, Y. Yao, B. Y. Zhao, and H. Zheng, "Object recognition and navigation using a single networking device," in *Proc. 15th Annu. Int. Conf. Mobile Syst. Appl. Serv. (MobiSys)*, 2017, pp. 265–277.
[48] W. Jiang *et al.*, "Towards 3D human pose construction using WiFi," in *Proc. 26th Annu. Int. Conf. Mobile Comput. Netw.*, 2020, pp. 1–14.
[49] F. Wang, S. Zhou, S. Panev, J. Han, and D. Huang, "Person-in-WiFi: Fine-grained person perception using WiFi," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 5452–5461.
[50] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 7291–7299.
[51] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "RMPE: Regional multi-person pose estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2334–2343.
[52] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2961–2969.

[53] K. Kolev, T. Brox, and D. Cremers, "Fast joint estimation of silhouettes and dense 3D geometry from multiple images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 493–505, Mar. 2012.
[54] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 834–849.
[55] L. Sigal, A. O. Balan, and M. J. Black, "HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *Int. J. Comput. Vis.*, vol. 87, nos. 1–2, p. 4, 2010.
[56] J. F. Federici, D. Gary, R. Barat, and D. Zimdars, "THz standoff detection and imaging of explosives and weapons," in *Proc. Opt. Photon. Global Homeland Security*, vol. 5781, 2005, pp. 75–84.
[57] R. M. Woodward *et al.*, "Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue," *Phys. Med. Biol.*, vol. 47, no. 21, 2002, Art. no. 3853.
[58] B. Allen, B. Curless, B. Curless, and Z. Popović, "The space of human body shapes: Reconstruction and parameterization from range scans," *ACM Trans. Graph.*, vol. 22, pp. 587–594, Jul. 2003.
[59] K. B. Cooper *et al.*, "Penetrating 3-D imaging at 4- and 25-m range using a submillimeter-wave radar," *IEEE Trans. Microw. Theory Techn.*, vol. 56, no. 12, pp. 2771–2778, Dec. 2008.
[60] D. M. Sheen, D. L. McMakin, and T. E. Hall, "Three-dimensional millimeter-wave imaging for concealed weapon detection," *IEEE Trans. Microw. Theory Techn.*, vol. 49, no. 9, pp. 1581–1592, Sep. 2001.

**Feng Zhang** (Member, IEEE) received the B.S. and M.S. degrees from the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China, in 2011 and 2014, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA, in December 2018.

He was with Origin Wireless Inc., Greenbelt, MD, USA, from December 2018 to May 2020. He is currently with Amazon.com Inc., Seattle, WA, USA. His research interests include wireless sensing, statistical signal processing, and wireless indoor localization.

Dr. Zhang was a recipient of 2020 IEEE INTERNET OF THINGS JOURNAL best paper award.

**Chenshu Wu** (Member, IEEE) received the B.S. degree from the School of Software, Tsinghua University, Beijing, China, in 2010, and the Ph.D. degree from the Department of Computer Science, Tsinghua University in 2015.

He is currently an Assistant Research Scientist with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA. He is also the Chief Scientist with Origin Wireless Inc., Greenbelt, MD, USA. His research interests include Internet of Things, wireless sensing, ubiquitous computing, and mobile health.

Dr. Wu is a Member of the ACM.

**Beibei Wang** (Senior Member, IEEE) received the B.S. degree (Hons.) in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2004, and the Ph.D. degree in electrical engineering from the University of Maryland, College Park, MD, USA, in 2009.

She was with the University of Maryland as a Research Associate from 2009 to 2010, and Qualcomm Research and Development from 2010 to 2014. Since 2015, she has been with Origin Wireless Inc., Greenbelt, MD, USA, where she is currently the Vice President of Research. She is also affiliated with the University of Maryland. She has coauthored of *Cognitive Radio Networking and Security: A Game-Theoretic View* (Cambridge University Press, 2010), and *Wireless AI: Wireless Sensing, Positioning, IoT, and Communications* (Cambridge University Press, 2019). Her research interests include Internet of Things, mobile computing, wireless sensing and positioning, and communications and networking.

Dr. Wang was a recipient of the 2020 IEEE INTERNET OF THINGS JOURNAL Best Paper Award, the 2015 IEEE Signal Processing Society Overview Paper Award, and the several research and invention awards from the University of Maryland. She has served on the Editorial Board of IEEE SIGNAL PROCESSING LETTERS, IEEE INTERNET OF THINGS JOURNAL, and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.

**K. J. Ray Liu** (Fellow, IEEE) is a Distinguished University Professor and a Distinguished Scholar-Teach with the University of Maryland, College Park, MD, USA, where he is also the Christine Kim Eminent Professor of Information Technology. He leads the Maryland Signals and Information Group conducting research encompassing broad areas of information and communications technology with recent focus on wireless AI for indoor tracking and wireless sensing.

Dr. Liu was a recipient of two IEEE Technical Field Awards: the 2021 IEEE Fourier Award for Signal Processing and the 2016 IEEE Leon K. Kirchmayer Graduate Teaching Award, the IEEE Signal Processing Society 2009 Technical Achievement Award, the IEEE Signal Processing Society 2014 Society Award, and the over a dozen of best paper/invention awards. As the founder of Origin Wireless, his invention won the 2017 CEATEC Grand Prix and CES 2020 Innovation Award. He also received the teaching and research recognitions from University of Maryland, including university-level Invention of the Year Award; and the College-Level Poole and Kent Senior Faculty Teaching Award, the Outstanding Faculty Research Award, and the Outstanding Faculty Service Award, all from A. James Clark School of Engineering. Recognized by Web of Science as a Highly Cited Researcher. He is 2021 IEEE President-Elect. He was the IEEE Vice President, the Technical Activities, and a Member of IEEE Board of Director as Division IX Director. He has also served as the President of IEEE Signal Processing Society, where he was the Vice President—Publications and an Editor-in-Chief of *IEEE Signal Processing Magazine*. He is a Fellow of AAAS and the U.S. National Academy of Inventors.