3D Point Cloud Generation with Millimeter-Wave Radar

KUN QIAN, University of California San Diego
ZHAOYUAN HE, University of California San Diego
XINYU ZHANG, University of California San Diego

Emerging autonomous driving systems require reliable perception of 3D surroundings. Unfortunately, current mainstream perception modalities, *i.e.*, camera and Lidar, are vulnerable under challenging lighting and weather conditions. On the other hand, despite their all-weather operations, today's vehicle Radars are limited to location and speed detection. In this paper, we introduce MilliPoint, a practical system that advances the Radar sensing capability to generate 3D point clouds. The key design principle of MilliPoint lies in enabling synthetic aperture radar (SAR) imaging on low-cost commodity vehicle Radars. To this end, MilliPoint models the relation between signal variations and Radar movement, and enables self-tracking of Radar at wavelength-scale precision, thus realize coherent spatial sampling. Furthermore, MilliPoint solves the unique problem of specular reflection, by properly focusing on the targets with post-imaging processing. It also exploits the Radar's built-in antenna array to estimate the height of reflecting points, and eventually generate 3D point clouds. We have implemented MilliPoint on a commodity vehicle Radar. Our evaluation results show that MilliPoint effectively combats motion errors and specular reflections, and can construct 3D point clouds with much higher density and resolution compared with the existing vehicle Radar solutions.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing**.

Additional Key Words and Phrases: FMCW Radar, SAR Imaging, Tracking, Radar Point Cloud

ACM Reference Format:
Kun Qian, Zhaoyuan He, and Xinyu Zhang. 2020. 3D Point Cloud Generation with Millimeter-Wave Radar. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 148 (December 2020), 23 pages. https://doi.org/10.1145/3432221

1 INTRODUCTION

Recent years have witnessed a surging demand for autonomous driving, driven by which the automotive industry revenue will expand by 30% to about $1.5 trillion, and 15% of new cars sold are expected to be fully autonomous by 2030 [15]. However, such predictions are optimistically based on the resolution of major technical issues. Currently, even the most advanced self-driving systems are still conditional automation at the Level 3 [25, 57], *i.e.*, allowing driver "eyes-off" most of the time but with occasional intervention. One major challenge in the way is the accidental failure of system perception, which reduces safety factors and even causes fatalities [61, 62].

The perception function relies on sensors that robustly capture key information of surroundings, such as nearby vehicles, pedestrians, and lanes. A wide range of sensing modalities is already available on vehicles, such as Lidar and camera. However, these sensors are limited by their operating medium and may not function well under certain conditions. Specifically, Lidar relies on the projection of laser beams that cannot penetrate opaque obstacles and are vulnerable to failure in harsh weather conditions. Cameras passively capture light scattered by

Authors' addresses: Kun Qian, University of California San Diego, qiank10@gmail.com; Zhaoyuan He, University of California San Diego, zhh159@eng.ucsd.edu; Xinyu Zhang, University of California San Diego, xyzhang@ucsd.edu.

2474-9567/2020/12-ART148 $15.00
https://doi.org/10.1145/3432221

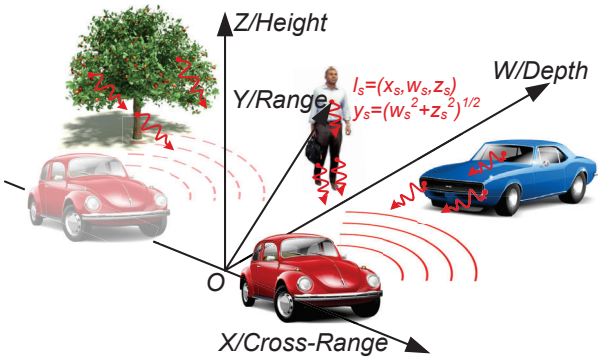


Fig. 1. Usage scenario and coordinate system in MilliPoint. The MilliPoint Radar moves along the cross-range direction and coherently combines the signals across locations to generate 3D point clouds.

objects, and cannot work in dark environments. In contrast, Radar actively transmits RF signals and processes the reflections to sense nearby objects' locations and moving speed. It works even under bad weather and poor lighting. However, vehicle Radar has two main drawbacks. First, it usually has a small form factor that limits the number of antennas (just like the number of pixels on a camera sensor), leading to low spatial resolution. While mechanical scanning Radar may achieve finer resolution, they have a large form factor due to the use of rotating horn antennas. The high-cost hardware also hinders wide adoption. For example the CTS 350X Radar [44] has a resolution of 0.9^o but costs around $19,000, almost as expensive as a midsize car. Second, vehicle Radars typically emit millimeter wave (mmWave) signals, which are specularly reflected off the surfaces of most objects and may not even be captured by the Radar. Therefore, although state-of-the-art wideband multi-antenna Radar can sense the locations of reflection points, the resulting "point cloud" tend to be extremely sparse (*e.g.*, a single point on a flat surface) [4, 21], which obviously cannot satisfy the feature-rich perception functions such as scene segmentation and object tracking.

Recent innovations in mmWave sensing have explored signal enhancement solutions such as fusing the measurements along the Radar's moving trajectory. For example, in RSA [70], a static 802.11ad mmWave transmitter illuminates the monitoring area, whereas a receiver moves and continuously measures angle of arrival (AoA) and received signal strength (RSS) from strong reflectors, which are combined to estimate locations, curvatures, and boundaries of close-by targets. Ulysses [69] moves co-located transmitter and receiver and uses steerable phased array beams to measure the angles and RSS of reflecting points to reconstruct the targets' surfaces. These approaches process RF signals received along the moving trajectory separately and then combine the derived object parameters (*e.g.*, AoA, RSS) to image objects, which we term as *non-coherent imaging*. While to some extent alleviating the specular reflection problem by illuminating the target from diverse locations, their imaging resolution is still limited by the physical size of the antenna array and cannot be fundamentally improved through spatial sampling. Furthermore, these approaches are only designed to sense the 2D location and coarse outline of close-by targets, whereas 3D Lidar-like perception is needed for reliable real-world autonomous driving applications. Today's most advanced radar-driven autonomous vehicles, such as Bertha from Benz [10], mainly rely on the Doppler properties of a limited number of reflection points on targets, with occasional camera intervention, to detect the objects. Nonetheless, a standalone Radar-based framework is still needed for all-weather perception, and ideally, it should approach the Lidar resolution.

It has been well established that the spatial resolution of Radar sensing is proportional to the *antenna aperture*, *i.e.*, number of elements along each dimension of the Radar's antenna plane (assuming fixed inter-element spacing) [38]. Based on this principle, synthetic aperture Radar (SAR) imaging algorithms synthesize a large antenna aperture by moving the Radar, and coherently combining the RF signals at different locations, as if there are

many virtual antenna elements along the way [6]. In what follows, we term SAR imaging as *coherent imaging* and use them interchangeably. While SAR imaging has become mature and widely used in areas such as remote sensing [29, 68] and security checkpoint [27, 50], directly applying SAR imaging to vehicle Radar poses three major technical challenges. *First*, coherent signal combining requires precise localization of the Radar device at the scale of the signal wavelength which is only several millimeters for vehicle Radar. Such precision, however, is not achievable in existing mobile localization schemes. *Second*, correct SAR imaging requires a proper *focusing* mechanism which, intuitively, defines the "aiming direction" of the SAR and the central coordinate of the resulting image. A straightforward method is to set the image center on the bisector of the antenna aperture, just like the center of a camera image lies in the lens' pointing direction. Such a method has been used in SAR-based remote sensing [6]. However, due to specular reflections, the Radar is "blinded" and receives no signals at certain locations along its moving trajectory. So directly focusing at the middle of its aperture may result in a blank image with vanished objects. *Third*, SAR can only improve the resolution along *cross-range dimension* (*i.e.*, the Radar's moving direction) which, combined with *range dimension* (perpendicular to the cross-range), forms a 2D image plane. However, the range dimension encodes not only the *depth*, but also the crucial *height information*. Without extracting the height, it is impossible to discriminate the target's size and shape.

To overcome these challenges, we propose MilliPoint, an automotive Radar system that aims to generate dense, high-resolution 3D point clouds of surrounding objects. As shown in Fig. 1, MilliPoint exploits natural linear motion of vehicle Radar and coherently combines the RF samples along its trajectory to improve sensing resolution. To accurately locate the Radar along the cross-range, our key observation is that different Tx/Rx antenna pairs on the Radar may experience similar channel responses with a lagging effect, and the delay depends on the spacing between antenna pairs. MilliPoint thus continuously estimates the delay, and translates it into relative location of the Radar along the cross-range aperture, with *millimeter level precision*. To overcome the focusing artifacts, we first model the effective antenna aperture of the Radar taking into account the specular reflection effects. Inspired by light-field cameras [2], we design an *automatic focusing* algorithm that post-focuses on each object separately in the scene, and then synthesizes a multi-focused image. To generate 3D point clouds, MilliPoint exploits the limited antenna aperture along the vertical direction. For the sake of computation efficiency and height resolution, it takes as input the imaging results of individual Tx/Rx antennas, and then extracts height information based on the correspondence between different images' pixel values, much like 3D reconstruction from multi-view camera images [34].

We have prototyped MilliPoint on an off-the-shelf mmWave Radar and conducted extensive field experiments to verify its performance, in comparison with two existing approaches: static radar, and mobile Radar with non-coherent combination of samples [69]. We find that the MilliPoint Radar can accurately self-track, with a cumulative error of only 1.2%, which is comparable to a commercial stereo camera. Compared with the non-coherent imaging approach, MilliPoint generates much sharper images, with 5 dB higher SNR on average, enabling easier image segmentation and semantic processing. Owing to the automatic multi-focusing, MilliPoint can image specular objects even with 60° orientation deviation. Our field tests further show that MilliPoint can produce dense and high-resolution 3D point clouds in realistic road scenarios. It can also estimate reflectivity of target points, as an additional dimension of information to facilitate object perception.

To our knowledge, MilliPoint represents the first system to enable high-resolution high-density 3D point cloud generation on low-end vehicle Radar. Our core contributions are three folds. *First*, we propose a novel algorithm to perform simultaneous imaging and Radar self-tracking, leveraging the short-term channel correlation on the Radar antennas to achieve millimeter-level location precision. *Second*, we design an automatic multi-focusing scheme to overcome the effect of specular reflection, leading to the dense and precise estimation of reflecting points. We further extrapolate height information from multiple SAR images and eventually generate a 3D point cloud of the environment. *Third*, we implement and verify the MilliPoint design on an automotive radar, and conduct case studies in realistic driving scenes. We envision MilliPoint as a new type of sensor fusion

modality. With a field-of-view spanned by the range and cross-range direction, the MilliPoint point cloud can be post-processed to facilitate parking, lane change, blind spot detection, and other perception functions such as cross-traffic monitoring [71].

2 PRELIMINARY

This section introduces the fundamentals of SAR imaging, and challenges of using SAR to generate 3D point clouds in vehicular settings.

2.1 SAR Imaging Fundamentals

A typical vehicle Radar periodically transmits frequency modulated continuous wave (FMCW) pulses [38]. The frequency of FMCW signal increases linearly at a preconfigured rate during each pulse period. By measuring the difference of instantaneous frequencies between the received signal and transmitted signal, the time of flight (ToF) between the Radar and the reflecting point can be obtained, which easily translates into the distance.

On this basis, a classical SAR system further moves the FMCW Radar along a straight (cross-range) path to form a large virtual antenna array, to improve spatial resolution along the *cross-range* direction. The Radar signal is represented at two time scales. The time within the duration of a pulse is referred to as fast time t, and the timestamps when generating each pulse is slow time u. As shown in the coordinate system in Fig. 1, suppose the origin lies at the center of the entire virtual antenna aperture, and x and y axes correspond to the cross-range and range directions respectively. *SAR essentially locates the strong reflecting points within the 2D x-y plane.* SAR imaging regards objects as composed of scatter points. If a point scatter in the x-y plane is at $\vec{l}_s = (x_s, y_s)$, then the received FMCW signal is [6]:

$$s(u,t) = a_s \text{rect}\left(\frac{uv}{L}\right) \text{rect}\left(\frac{t}{T_p}\right) e^{-j\frac{4\pi(f_c+\gamma t)}{c} r_s(u)}, \tag{1}$$

where a_s is the signal amplitude, v is the moving velocity, L is the cross-range aperture size, T_p is the period of one pulse, γ is the linear increasing rate of the signal frequency, f_c is the center frequency of the pulse and c is the speed of light. The location of the Radar can be denoted as $\vec{l}_r = (uv, 0)$, and $r_s = \|\vec{l}_r - \vec{l}_s\|$ is the distance between the scatter and the Radar. The phase shift $\frac{4\pi(f_c+\gamma t)}{c} r_s(u)$ represents the round-trip propagation delay between the Radar and the scatter. rect$(\cdot)$ is the rectangular function, which equals 1 within $[0, 1]$ and 0 otherwise. The term rect $\left(\frac{uv}{L}\right)$ rect $\left(\frac{t}{T_p}\right)$ means that the Radar only receives signals within the cross-range aperture and during the period of each pulse.

Let $x = uv$, SAR intermediately applies a 1D FFT to transform the received signal from the x-t time domain to the k_x-k_r spatial frequency domain $S(k_x, k_r) = \mathbf{FFT}_{\mathbf{x}}[s(x,t)]$, where $k_x = -\frac{4\pi f_c}{c}\frac{x}{r_s(u)}$ and $k_r = \frac{4\pi(f_c+\gamma t)}{c}$. It should be noted that FFT requires that the x samples are equally spaced, *i.e., the spatial intervals between FMCW pulses must be uniform.* It then replaces k_r with $k_y = (k_r^2 - k_x^2)^{\frac{1}{2}}$ to obtain $S(k_x, k_y)$. In effect, k_x, k_y, k_r are *spatial frequencies* transformed from the cross-range x, range y and distance r, respectively. $s(u,t)$ has slowly varying amplitude and rapidly varying phase over u. Thus, when applying the integration of the 1D FFT over u, most $s(u,t)$ whose phase varies rapidly tend to cancel each other and only those stationary points with zero phase derivative remains. Thus, $S(k_x, k_y)$ can be approximated by $s(u,t)$ at the stationary points where the derivative of their phases is zero, which gives [6]:

$$\begin{aligned} |S(k_x,k_y)| &\approx a_s \text{rect}\left(\frac{k_r - 4\pi f_c/c}{4\pi\gamma T_p/c}\right) \text{rect}\left(\frac{k_x y_s - k_y x_s}{L k_y}\right) \\ \angle S(k_x,k_y) &\approx -k_x x_s - k_y y_s \end{aligned} \tag{2}$$

The rect$(\cdot)$ functions limit the non-zero cross-range frequency support of the scatter, *i.e.*, $\frac{x_s - \frac{L}{2}}{\sqrt{(x_s-\frac{L}{2})^2 + y_s^2}} k_r \le k_x \le$

$\frac{x_s+\frac{L}{2}}{\sqrt{(x_s+\frac{L}{2})^2+y_s^2}}k_r$, centering at $k_x=\frac{x_s}{\sqrt{x_s^2+y_s^2}}k_r$. $S(k_x,k_y)$ is then multiplied with a matched phase filter $\Phi(k_x,k_y)=e^{j(k_x x_s+k_y y_s)}$ to *focus on the scatter point*. That is, by compensating the phase shift in the spatial frequency domain, the scatter is shifted to the image center $(0,0)$ in the space domain. It is noted that, in classical SAR, *the coarse location of the scatter must be known a priori for generating the matched phase filter and imaging with the highest quality*. Finally, the non-zero frequency support around the center $k_x=\frac{x_s}{\sqrt{x_s^2+y_s^2}}k_r$ is selected and a 2D IFFT is applied to generate image point $f(x,y)=\mathbf{IFFT}_{\mathbf{k_x},\mathbf{k_y}}[S(k_x,k_y)]$, which represents the reflecting intensity at location (x,y). In practice, a scatter within the aperture (*i.e.*, $-\frac{L}{2}\le x_s\le\frac{L}{2}$) always has non-zero support around $k_x=0$. It is thus acceptable to assign the image center on the bisection line of the cross-range aperture (*i.e.*, $x=0$) and select the non-zero support around $k_x=0$, although with a slight sacrifice of the image quality. However, we will show later (Sec. 2.2) that this classical focusing mechanism substantially reduces the visibility of specular reflectors when applied to automobile perception.

The SAR imaging algorithm can be directly extended to multiple scatters, as all operations involved are additive. In effect, for any object comprised of many scattering points, the algorithm output is the superimposition of all image points.

2.2 Challenges for Radar Point Cloud Generation

SAR imaging has been widely used in remote sensing and security check. In the former case, orbital Radar tracks itself via high-end GPS and inertial measurement unit (IMU) for motion error compensation and generates 2D images of rough terrains. Flat surfaces (*e.g.*, lake, road, airport) appear as empty spots due to their specular reflections. In the latter case, a massive phased array moves along a predefined track that fully encloses the target, and generates corresponding 3D images. It may be tempting to consider applying these existing SAR solutions to automobile Radar. However, three fundamental challenges have hindered such adaptation.

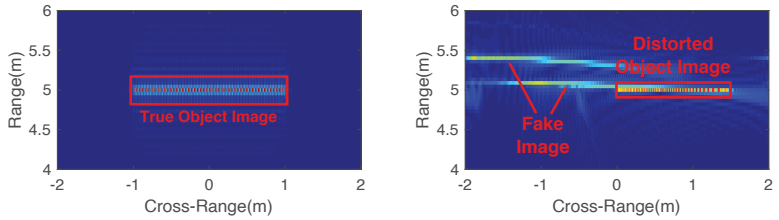
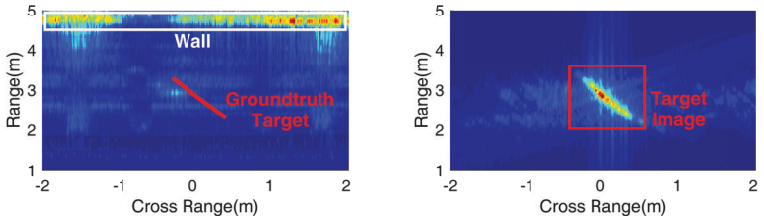


Fig. 2. Impact of aperture motion on SAR imaging. (a) Imaging result with uniform motion. (b) Imaging result with variable motion.

Aperture motion error. As stated in Sec. 2.1, the Fourier transform in SAR imaging requires linear aperture motion and uniform intervals between transceptions of pulses. While the former requirement can be guaranteed with millimeter-level accuracy for vehicle driving thanks to wheel alignment mechanisms [14], the latter can hardly be fulfilled due to speed variations. Consequently, the Radar samples occur at irregular intervals (corresponding to locations of the virtual antennas) along the vehicle's cross-range trajectory. To understand the impacts of such irregular aperture motion, we simulate both uniform and variable motion in a scenario with a 2 m wide linear object consisting of 40 scatter points with 5 cm spacing, as shown in Fig. 2a. In the case of variable motion, a Radar increases its moving speed from 0 with the acceleration of 2 m/s^2. We then apply the classical SAR imaging to both cases. Fig. 2b shows the resulting SAR image with variable motion. The image is distorted, showing erroneous location and width of the target, while introducing ghost images where no object exists.

Blindness by specular reflection. The wavelength of mmWave Radar signals is larger than the surface

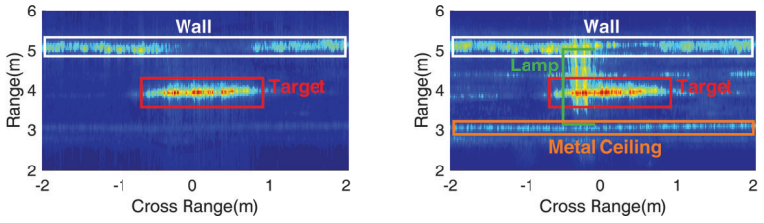


(a) (b)

Fig. 3. Impact of specular reflection. (a) Imaging result with wrong focusing (0^o). (b) Imaging result with correct focusing (45^o)

roughness of most objects on road. Thus, the Radar signals are specularly reflected, and can only be received by the Radar when it is around the normal directions of the specular surfaces. Therefore, for a specular object, its *effective aperture* is limited, even if the Radar moves a long way to create a large *physical aperture*. Said differently, the target may be beyond the coverage of the effective aperture while still within the physical aperture. By wrongly focusing on the bisection line of the physical aperture as in classical SAR, the non-zero support of the target will deviate from the image center. As a result, like the defocusing effect of a light-field camera, the Radar may generate a blurry image of the object, or even fail to sense the object and becomes "blind".

To understand the impact of specular reflection, we place a flat metal board in front of a TI Radar, its surfacing being 45° relative to the Radar cross-range direction. The ground-truth location of the board is marked with the red line segment in Fig. 3a. While the Radar is physically moved from -4 m to 4 m along cross-range, it can only receive specular reflections of the board from -4 m to -2 m, which constitutes the *effective aperture* of the board. However, by mistakenly focusing on the center of the physical aperture (*i.e.*, at 0 m), the target vanishes in the image and becomes almost invisible, as shown in Fig. 3a. In contrast, suppose the location and orientation of the specular target relative to its effective aperture is known, by focusing on the center of the target, it is evident that the target shape, location, and orientation can be reconstructed, as shown in Fig. 3b. In practical scenarios, however, the target's location is unknown prior to imaging. Moreover, multiple targets may be dispersed at different locations, possibly with different orientations, making it impossible to image all targets by focusing on only one imaging center. Due to safety concerns, the specular reflection problem is more critical for autonomous driving than conventional remote sensing applications.



(a) (b)

Fig. 4. Applying conventional SAR imaging to (a) a 2D scene with objects on the same plane and (b) a 3D scene with multiple objects at different heights.

Lack of height information for 3D point cloud generation. In addition to 2D locations, heights of scatter points are essential information for safety risk assessment, and for advanced perception functions. However, with 1D cross-range aperture, SAR can only project a 3D scene onto the 2D x-y plane. All scatters with the same cross-range and range values are stacked into the same pixel on the 2D image. For example, Fig. 4 compares SAR

imaging results for 2D and 3D indoor scenes. In the 2D environment, only objects on the same horizontal plane as the Radar are considered. In the 3D environment, objects at ceiling height such as lamp fixtures also superimpose on the image, defying segmentation algorithms. To image objects along the height direction, a vertical antenna array with similar dimensions as the objects is necessary, just like that used in airport security checkpoint [38]. However, this does not fit on a vehicle Radar. An alternative sensing framework is thus required to enable 3D point clouds.

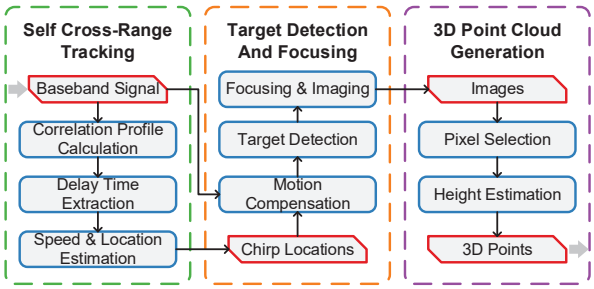


Fig. 5. System overview of MilliPoint.

3 SYSTEM OVERVIEW

MilliPoint aims to redesign SAR imaging to address its limitations on vehicle Radar and extend it to generate 3D point clouds. As shown in Fig. 5, MilliPoint consists of three major components, *self cross-range tracking*, *target detection and focusing*, and *3D point cloud generation*.

MilliPoint takes the Radar's baseband samples as the sole input. Upon receiving the samples, MilliPoint first tracks the Radar's relative movement with the *self cross-range tracking* module. Briefly speaking, it calculates the cross correlation between the received signals of two pairs of Tx/Rx antennas, from which it computes the time lag that leads to maximum correlation. Based on prior knowledge of the antenna spacing, it converts the time lag into instantaneous cross-range speed and then the location of the Radar (relative to the starting point of the virtual aperture).

Both the RF signal samples and corresponding location samples are then used in the *target detection and focusing* module to image objects in the 2D x-y plane. To overcome the effect of specular reflection, MilliPoint transforms RF data to the 2D cross-range frequency and range domain, where prominent scatters are detected and localized. The centers of clusters of prominent scatters are then used to guide an *automatic multi-focusing* mechanism which prevents the aforementioned blindness problem (Sec. 2.2). Finally, note that multiple pairs of Tx/Rx antennas along the vertical direction can create multiple images for the same scatter point. MilliPoint extracts the height of the point based on phase differences of the image copies. It then fuses the height with the x-y image plane to form a 3D point cloud, and denoises it and prepares it for post-processing algorithms.

4 SYSTEM DESIGN

We now introduce the model and detailed design of the three key components in MilliPoint.

4.1 Self Cross-Range Tracking

SAR imaging requires the cross-range spacing between consecutive pulse sampling locations be within $\frac{\lambda}{4}$, λ being signal wavelength, to avoid aliasing effect [6]. Equivalently, the inter-element spacing of the virtual antenna array should fall in $\frac{\lambda}{4}$, and the sampling positions must be tracked with millimeter level accuracy finer than $\frac{\lambda}{4}$. *Such tracking accuracy and update rate are not attainable using existing on-vehicle sensors, such as GPS and IMUs.* Thus, we seek the possibility of enabling Radar self-tracking by exploiting the motion cues hidden in RF signals.

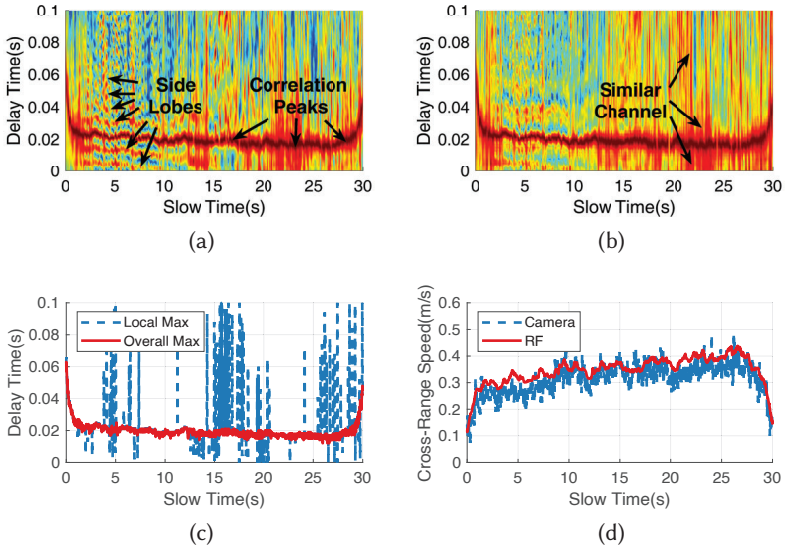


Fig. 6. Illustration of self cross-range tracking. (a) Cross correlation of single antenna pair. (b) Average cross correlation of multiple antenna pairs. (c) Extraction of delay time with maximum correlation. (d) Cross-range speed estimated by RF-based and camera-based self tracking.

Cross correlation of Radar samples. A mmWave Radar usually comprises multiple Tx and Rx *antenna pairs* (sometimes the Tx and Rx share the same antenna). Our key observation in the self-tracking design is that different antenna pairs may experience similar channel responses while moving along the cross-range, but with some delay associated with antenna spacing. This is because the real-world scenes tend to remain relatively static as the antenna pairs sequentially pass it at vehicle speed. Generally, suppose the current Radar location is $\vec{l}_R = (x_R, 0, 0)$; the relative locations of the i-th Tx/Rx antennas are $\vec{l}_t^{(i)} = (x_t^{(i)}, 0, 0)$ and $\vec{l}_r^{(i)} = (x_r^{(i)}, 0, 0)$, respectively. There are N scatters in the scene where the location of the n-th scatter is $\vec{l}_n = (x_n, y_n, z_n)$, and the Radar movement is characterized as $\vec{\delta}(\delta_u) = (\delta_x, \delta_y, \delta_z)$. Then the reception of the i-th antenna pair after movement $\vec{\delta}$ is:

$$s^{(i)}(u + \delta_u, t) = \sum_{n=1}^{N} a_n e^{-j\frac{2\pi(f_c + \gamma t)}{c} r_n(u+\delta_u)} \tag{3}$$

where $r_n(u + \delta_u) = \|\vec{l}_n - (\vec{l}_t^{(i)} + \vec{\delta})\| + \|\vec{l}_n - (\vec{l}_r^{(i)} + \vec{\delta})\|$ is the instantaneous range of the n-th scatter.

By correlating the received signals of two antenna pairs before and after the movement, we have:

$$\begin{aligned} C(u, \delta_u) &= \int_{-\frac{T_p}{2}}^{\frac{T_p}{2}} s^{(2)}(u + \delta_u, t) s^{(1)}(u, t)^* \mathrm{d}t \\ &= \sum_{n,l} a_n a_l e^{-j\frac{2\pi f_c}{c}\delta_r^{(n,l)}(u,\delta_u)} \int_{-\frac{T_p}{2}}^{\frac{T_p}{2}} e^{-j\frac{2\pi\gamma t}{c}\delta_r^{(n,l)}(u,\delta_u)} \mathrm{d}t \\ &= \sum_{n,l} a_n a_l e^{-j\frac{2\pi f_c}{c}\delta_r^{(n,l)}(u,\delta_u)} T_p \,\mathrm{sinc}\left(\frac{\pi B}{c}\delta_r^{(n,l)}(u, \delta_u)\right) \end{aligned} \tag{4}$$

where $\delta_r^{(n,l)}(u, \delta_u) = r_n(u + \delta_u) - r_l(u)$ is the range difference of the two scatters and B is the bandwidth of the

FMCW signal. On one hand, in terms where $n = l$, $\delta_r^{(n,n)}(u, \delta_u)$ can be approximated as:

$$\delta_r^{(n,n)}(u, \delta_u) \approx \frac{(\frac{\delta_p}{2} - \delta_x)x_n - \delta_y y_n - \delta_z z_n}{r_n} \tag{5}$$

where $\delta_p = (x_t^{(1)} - x_t^{(2)}) + (x_r^{(1)} - x_r^{(2)})$ is the spacing of two antenna pairs. Given that the vehicle moves along cross-range direction, *i.e.*, $\delta_x \gg \delta_y, \delta_z$, the terms $\delta_y y_n$ and $\delta_z z_n$ can be neglected. As a result, the terms with $n = l$ reach the maximum when cross-range movement $\delta_x = \frac{\delta_p}{2}$. On the other hand, for terms where $n \neq l$, $\delta_r^{(n,l)}(u, \delta_u)$ tends to be large, and the sinc function output is small and likely to cancel each other, resulting in little contribution of these terms. In short, *the correlation $C(u, \delta_u)$ reaches the maximum when $\delta_x = \frac{\delta_p}{2}$*.

We further verify this model with a field test, where a TI FMCW Radar moves along cross-range for about 30 s. The Radar has 6 Tx and 8 Rx antennas (Fig. 9b shows its layout), and 1 kHz pulse repetition rate. For brevity, we denote an antenna pair as $\langle i, j \rangle$, representing that the antenna pair consists of the i-th Tx antenna and the j-th Rx antenna. We select two antenna pairs, $\langle 2, 4 \rangle$ and $\langle 6, 1 \rangle$. The two Tx antennas are 3.5λ apart and the two Rx antennas have the same azimuth location, resulting in $\delta_x = 1.75\lambda$. The received signal of each FMCW pulse is correlated with the last 100 pulses. Fig. 6a shows the corresponding correlation profile. A sequence of correlation peaks with consistent delay time is identified as the Radar moves, which matches the foregoing model.

Robust self-tracking. Given that the relative locations of the Tx/Rx antennas $x_t^{(i)}$ and $x_r^{(i)}$ are known *a priori*, if the corresponding delay time δ_u can be correctly extracted from the correlation profile, the Radar's cross-range moving speed should be: $v_x = \frac{\delta_x}{\delta_u}$. The relative cross-range location of the Radar can be obtained by further integrating the speed. In practice, however, directly selecting sequences of maximum correlation values is error-prone due to the corruption of interferences, as evident from Fig. 6a. Instead, we identify two main types of interferences and develop corresponding sanitizing steps to robustly extract the delay time.

First, besides the global maximum at δ_x, the correlation $C(u, \delta_u)$ also reaches the local maximum at side lobe peaks of the sinc function. In cases with very few scatters in the scene, the side lobes of sinc components in Eq. (4) are less likely to cancel out each other. For example, they are observable within 3-10 s in Fig. 6a. Due to noises and the approximation in Eq. (5), these side peaks may become higher than the main peak, which misleads the delay time estimation. To overcome the side peaks, MilliPoint exploits multiple antenna pairs with different spacing. Specifically, MilliPoint assumes that the Radar moves at a constant speed within each short delay time interval (e.g., 0.1 s). Thus, it scales each correlation profile along the delay time dimension, by virtually changing the spacing of the corresponding antenna pair to a reference distance, *e.g.*, the spacing of the 1st antenna pair, to align the peaks of the maximum correlation values in all correlation profiles. Then, the average of all scaled correlation profiles is computed. The scaling and averaging can be formulated as:

$$\hat{C}(u, \delta_u) = \frac{1}{M} \sum_{i=1}^{M} C_i(u, \frac{\delta_{p_i}}{\delta_{p_1}} \delta_u) \tag{6}$$

where M is the number of antenna pairs, C_i and δ_{p_i} are the correlation profile and the spacing of the i-th antenna pair. With this operation, the main peaks of all antenna pairs reinforce each other, while side peaks are averaged out. To verify this mechanism, we repeat the foregoing experiments, but add the correlation profiles of two additional antenna pairs, *i.e.*, $\langle 4, 3 \rangle$ and $\langle 2, 4 \rangle$ and $\langle 5, 2 \rangle$ and $\langle 2, 4 \rangle$. It is noted that no new antenna hardware is needed during the averaging process. Fig. 6b shows the resulting correlation profile of the three antenna pairs, where the side lobes of sinc components are significantly reduced.

Second, when the Radar passes by a large homogeneous reflector, it may continuously experience similar channel, resulting in large correlation values for a long period and erroneous maximum peaks due to noises, as evident in Fig. 6b from 13 s to 30 s. Instead of locally selecting maximum peaks, MilliPoint searches for

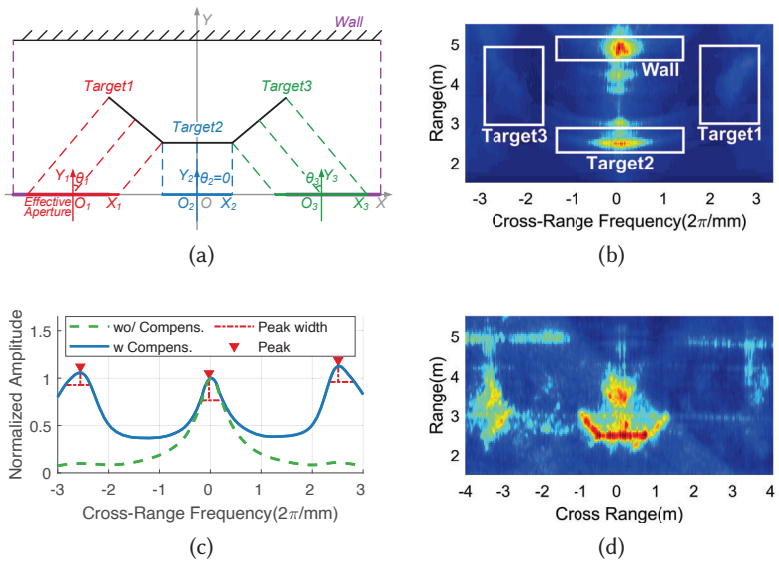


Fig. 7. Illustration of target detection and focusing. (a) Scene setup for model analysis and experimental validation. (b) Spectrum in 2D cross-range frequency and range domain. (c) Detection of targets in cross-range domain. (d) Combining image of partial results with different image centers.

continuous delay time sequence with maximum power, by constraining the change of delay time:

$$\begin{aligned} \delta_{u,m}(u) &= \arg\max_{\delta_u(u)} \hat{C}(u, \delta_u) \\ \text{s.t.} \forall u, &|\delta_u(u+\delta_u) - \delta_u(u)| \le \beta \end{aligned} \tag{7}$$

where δ_u is the time interval between adjacent pulses and β is the maximum change of delay time between adjacent pulses. The problem can be modeled as a dynamic programming problem and efficiently solved as in [43]. Fig. 6c illustrates this process, again based on the samples in the previous experiment. Fig. 6d further plots the corresponding moving speed estimation, and compares it with the ground-truth measurement from a stereo camera (a ZED camera [53] with sub mm of self-tracking precision). We see that the result of our self-tracking algorithm closely matches the ground-truth, demonstrating the feasibility of precise self-tracking using a standalone Radar.

4.2 Automatic Multi-Focusing

In this section, we first model the effect of the specular reflection problem (Sec. 2.2), and then develop an automatic multi-focusing mechanism to overcome it. For ease of exposition, we set up an example scene (Fig. 7a), where we place three metal boards with incident angles (135°, 0° and 45°) to the cross-range direction (x-axis). Background objects also exist, to represent sophisticated real 3D environment.

Modeling of Specular Reflection. Consider an arbitrary target n in the scene. Since the Tx and Rx antennas are closely co-located, the Radar can only receive reflections around the normal direction of the target n due to specular reflection. As a result, the approximate *effective aperture* corresponding to the target n is shortened and its center is shifted to O_n, as shown in Fig. 7a. When multiple objects exist in the scene, each will have its effective aperture, which is part of the physical aperture. Unfortunately, without knowing the location, orientation, and size of the target, it is impossible to determine the effective aperture which is part of the *physical aperture*, *i.e.*, the total distance that the Radar moves along the cross-range.

Suppose the midpoint and length of the unknown effective aperture of the target n are o_n and L_n respectively. The overall received signal $s(u,t)$ and cross-range frequency spectrum $S(k_x,k_r)$ can be modeled by summing over each target's contribution within its effective aperture:

$$\begin{aligned} s(u,t) &= \sum_{n=1}^{N} s_n(u-\frac{o_n}{v},t)\cdot \text{rect}\left(\frac{uv-o_n}{L_n}\right) \\ S(k_x,k_r) &\approx \sum_{n=1}^{N} S_n(k_x,k_r)e^{jk_x o_n} \end{aligned} \tag{8}$$

where $s_n(u,t)$ and $S_n(k_x,k_r)$ are the received signal and frequency spectrum of the target n, in the local coordinate centered at the midpoint of the its effective aperture. It means that the frequency response of the target n has the same coordinate in both $S(k_x,k_r)$ and $S_n(k_x,k_r)$. This implies that even without the knowledge of a target's effective aperture, it is still feasible to detect the target in $S(k_x,k_r)$, where the location of the target relative to its effective aperture can be derived for correct focusing and thus imaging.

Object Detection and Focusing. MilliPoint's automatic multi-focusing mechanism builds on the above insight of effective aperture to overcome the specular reflection. Inspired by light-field cameras [2, 39], MilliPoint post-focuses on each target in the scene separately, from which it synthesizes a full image.

To focus on and image any target n, MilliPoint must first estimate the target center $\vec{l}_{s_n}$ and aperture length L_n. According to Eq. (2), the target center can be approximately determined by the distance $\|\vec{l}_{s_n}\|$, and the incident angle θ_n between the norm of the target n and the y-axis (Fig. 7a):

$$\begin{aligned} x_{s_n} &= \|\vec{l}_{s_n}\|\sin\theta_n \\ y_{s_n} &= \|\vec{l}_{s_n}\|\cos\theta_n \end{aligned} \tag{9}$$

Recall in Sec. 2.1 that the center of the non-zero support of the n-th target $k_x = \frac{x_{s_n}}{\sqrt{x_{s_n}^2+y_{s_n}^2}}k_r = k_r\sin\theta_n$ is proportional to $\sin\theta_n$, a 1D IFFT is applied to $S(k_x,k_r)$ to obtain $\hat{S}(k_x,r) = \mathbf{IFFT}_{\mathbf{k_r}}[S(k_x,k_r)]$. MilliPoint then detects the targets with prominent support in $\hat{S}(k_x,r)$, and estimate their θ_n and $\|\vec{l}_{s_n}\|$ with the corresponding k_x and r.

For example, Fig. 7b shows the $\hat{S}(k_x,r)$ corresponding to the scene in Fig. 7a. The non-zero supports of main targets are marked with white boxes. To *automatically detect and localize* them, MilliPoint takes two steps of peak finding. *First*, the distance values estimated by the Radar are quantized into bins (resolution determined by the FMCW bandwidth). For each distance bin, MilliPoint finds peaks of the cross-range frequency spectrum that are higher than ϵ of the maximum peak, where $\epsilon\in[0,1)$ is an empirical threshold for rejecting false noisy peaks. All peaks are aggregated to generate an overall cross-range frequency spectrum $\bar{a}(k_x) = \sum_r I(k_x,r)\hat{S}(k_x,r)$, where $I(k_x,r)$ is the indicator of peaks. By keeping the peak values only, MilliPoint prevents the strong reflectors from obfuscating weak ones. *The peaks in $\bar{a}(k_x)$ are identified as targets' reflection points. The effective aperture lengths of targets are approximated by the corresponding peak widths.*

In practice, reflections of targets are affected by antenna directivity and target distance. To accurately detect targets, MilliPoint compensates these two factors in the spectrum $\hat{S}(k_x,r)$ before detecting peaks, *i.e.*, $\bar{S}(k_x,r) = \frac{\eta(r)}{\eta(\theta)}\hat{S}(k_x,r)$, where $\eta(\theta)$ is the antenna gain along θ which can be obtained in advance through a one-time measurement, and $\eta(r)\propto r$ is the attenuation over distance. Fig. 7c shows the overall cross-range spectrum. Note that with compensation, the peaks on sides become more salient. Three peaks are detected, corresponding to the three targets and wall, where the peaks of the target 2 and wall coincide.

Second, for each peak detected in the overall cross-range spectrum, MilliPoint further identifies the peak along the range direction. Then, it uses the 2D location (range, cross-range) of each peak as the image center, and applies the classical SAR imaging algorithm (Sec. 2) to image the corresponding object. All images are then combined by selecting the maximum value of each pixel across all images to synthesize the overall image. Fig. 7d

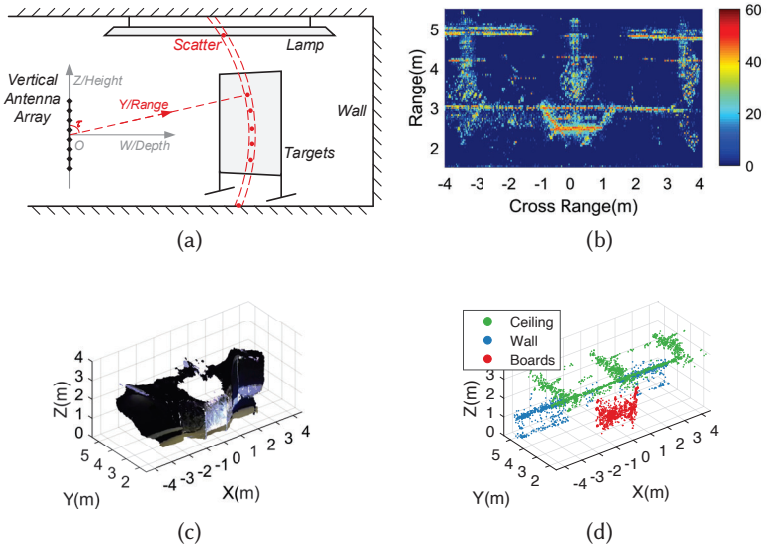


Fig. 8. Illustration of 3D point cloud generation. (a) Scene setup for model analysis and experimental validation. (b) Pixel selection based on neighbor similarity (c) 3D point clouds by ZED stereo camera. (d) 3D point clouds by MilliPoint.

shows the combined image of the scene in Fig. 7a, where all three metal boards and the unoccluded parts of the wall are correctly imaged.

4.3 3D Point Cloud Generation

Recall that, with 1D cross-range aperture generated by moving an antenna pair, SAR imaging can only generate a 2D image spanning the x-y dimensions. Objects with different heights (elevation angles) may all be stacked on the same x-y plane and obfuscate each other. Fortunately, though unable to image objects along the height direction due to its small aperture size, the vehicle Radar can still discriminate objects at different heights using multiple vertical Tx/Rx antennas. One intuitive way to leverage the vertical aperture is to digitally beamform to different elevation angles and then run SAR imaging. However, this method has two drawbacks. *First*, the number of SAR imaging operations is proportional to the number of elevation angles. A huge number of beam scans are needed to achieve an acceptable angular resolution (*e.g.*, 1°), hence unsuitable for time-critical high mobility scenarios. *Second*, even with a fine angular resolution, objects may have non-uniform height resolution, as resolution decreases as range increases. To overcome such barriers, MilliPoint applies SAR imaging to each Tx/Rx antenna pair, and *leverages co-registered pixels generated by different antenna pairs to estimate the height of the corresponding target scatter point.* It then forms a 3D point cloud by fusing height information with the x-y plane image.

Height estimation with co-registered pixels. To extract height from image pixels, we first model the phase responses of pixels. According to Eq. (2), the phase response of a point scatter at $\vec{l} = (x, y)$ after matched filtering and interpolation is:

$$\begin{aligned} \angle S(k_x, k_y) &= -k_x(x - x_s) - k_y(y - y_s) \\ k_x &\in [-\frac{\pi}{\delta_x}, \frac{\pi}{\delta_x}] \\ k_y &\in [k_{y_0} - \frac{2\pi f_c \gamma T_p}{c}, k_{y_0} + \frac{2\pi f_c \gamma T_p}{c}] \end{aligned} \tag{10}$$

where δ_x is the cross-range sample spacing, and $k_{y_0} = \frac{4\pi f_c}{c}\frac{y_s}{r_s}$ is the center spatial frequency of range dimension and depends on the imaging center. After IFFT, the phase response of the pixel corresponding to the scatter is:

$$\angle f(x, y) = -k_{y_0}(y - y_s). \tag{11}$$

where $k_{y_0} y_s$ is constant for all pixels, $k_{y_0} y$ encodes range information and can be used to estimate height and depth, as shown in Fig. 8a. On this basis, we adapt the idea of AoA estimation with antenna array [65] to the virtual pixel array. Specifically, suppose the height of the scatter is h, and its elevation AoA is $\tau = \text{acos}\,\frac{h}{y}$. We use an *array steering vector* $\vec{a}(h)$ to characterize the phases of the array of pixels relative to the first pixel, as a function of the height of the scatter:

$$\vec{a}(h) = (1, e^{-jk_{y_0}\delta_a\frac{h}{y}}, \cdots, e^{-jk_{y_0}(N-1)\delta_a\frac{h}{y}})^T \tag{12}$$

where δ_a is the physical antenna spacing and N is the number of antenna pairs. It means that the range difference between two adjacent pixels is $\delta_a \cos\tau = \delta_a\frac{h}{y}$, and the corresponding phase different is $k_{y_0}\delta_a\frac{h}{y}$. Then, MilliPoint estimates the height of the pixel using the Capon algorithm [5], which essentially finds the steering vector that intensifies the incident signal and minimizes the power contribution of signals from other directions:

$$h_s = \arg\min_h (\vec{a}(h)^H R_s^{-1} \vec{a}(h)) \tag{13}$$

where $R_s = \vec{f}_s \vec{f}_s^H$ is the covariance matrix and $\vec{f}_s$ is the vector of co-registered pixel values.

The depth is further calculated as $w_s = \sqrt{y_s^2 - h_s^2}$. On this basis, the scatter point corresponding to the pixel is localized in 3D space as (x_s, w_s, h_s). Furthermore, for each scatter, the signal strength of its reflected signal can be calculated by beamforming towards the scatter, where the beam steering vector is computed based on the estimated height, *i.e.*,

$$p_s = 1/N \cdot |\vec{a}(h_s)^H \vec{f}_s| \tag{14}$$

Selection of prominent pixels. In classical remote sensing and surveillance applications, SAR is installed on airplanes/satellites, with an overlooking view and most pixels corresponding to real scatters on the ground. In contrast, when applied to autonomous driving, SAR has a lateral view, where most signals propagate through empty space without reflections and the resultant images usually only contain a few clusters of pixels corresponding to real objects. Thus, it is necessary to process these pixels beforehand, in order to provide noiseless point clouds, and avoid wasting of computation on invalid pixels. One intuitive scheme is to select pixels with prominent amplitude. However, it is not straightforward to use a universal threshold, as objects with different materials and range may have different reflection strength. Instead, MilliPoint exploits the observation that prominent pixels are less like their neighboring pixels. Specifically, for each pixel, it calculates the variance of amplitudes of neighboring pixels. Then, it counts the number of neighbor pixels whose amplitudes are one standard deviation smaller. An overall threshold on the number of neighbor pixels is applied to select those prominent center pixels. Theoretically, the more neighboring pixels are selected, the more prominent the center pixel is, but the more likely a valid point is falsely filtered out. So the threshold can be adjusted to trade off the density of point clouds and the number of invalid points.

Furthermore, MilliPoint takes two denoising steps to mitigate the impact of random noises on pixel selection. *First*, it rearranges the 2D image generated by each antenna pair into a 1D vector. All such vectors are then stacked into a 2D matrix, upon which a principle component analysis is performed. The first principle component is selected and rearranged back into a 2D image to intensify meaningful pixels and filter others. *Second*, MilliPoint applies the TV-L1 model [7] to further reduce noise. Briefly speaking, TV-L1 finds an objective image closely matching the original image, but has the minimum total variance (TV), defined as the integral of the absolute gradient of the image, in order to remove noise in the original image. Fig. 8b shows an example of pixels selected corresponding to the real scene in Fig. 8a (the same as in Fig. 7a). Specifically, a neighbor area with 1 cm wide in cross-range and 22 cm long in range is set, resulting in a total number of 55 pixels. In the image, prominent pixels corresponding to real objects, *e.g.*, boards, lamps, and the wall, are highlighted and can be selected for

post-processing.

Processing of point clouds. Each point in the point cloud is represented by its 3D location (*i.e.*, cross-range, depth, and height) and amplitude of reflection. The point cloud can be further used as input for a wide range of automobile perception algorithms.

To showcase the usage of the 3D point cloud, we manually segment the point cloud into different objects. We use a ZED stereo camera (as in Fig. 9) to obtain the visual 3D point cloud of the scene in Fig. 8a. The three metal boards, three lamps on the ceiling, and the background wall are visible in the scene. Fig. 8d shows the segmentation result of the point cloud generated by MilliPoint. Points belonging to the same segment share the same color. While co-located in 2D cross-range and range domain, the prominent reflectors, including three boards, three lamps, and the unoccluded parts of the wall are localized and separated according to their height differences. In addition, a metal crossbeam, which is on the ceiling but out of the scope of the ZED camera, is also captured by the Radar, thanks to its wider angle of view. Further analysis and processing of point clouds are beyond the scope of this paper and left as future work.

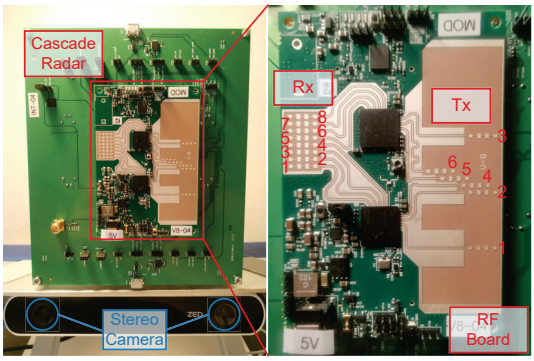


Fig. 9. Experimental testbed for verifying MilliPoint.

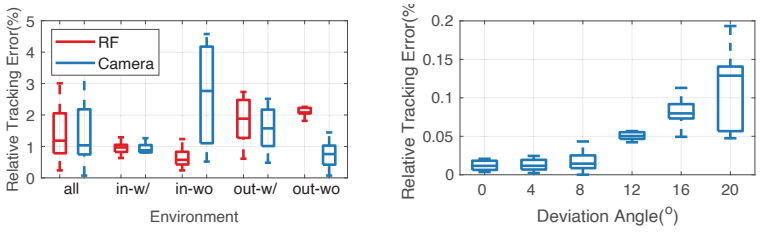
5 EXPERIMENTAL RESULTS

5.1 Experimental Setup

Implementation. We implement MilliPoint using a commercial millimeter wave Radar evaluation platform. As shown in Fig. 9, the platform consists of two sub-modules, a customized TI dual-chip FMCW Radar sensor with 6 Tx and 8 Rx antennas, and a TI MMWCAS-DSP-EVM board for acquiring baseband signal samples and streaming them to a PC host. By default, we set the FMCW parameters as follows: pulse duration 60 μs, frequency slope 66 MHz/μs, baseband sampling rate 5 Msps, and the number of samples acquired in each pulse is 256, which translates into a range resolution of 0.044 m and maximum range 11.31 m. The pulse repetition rate is set to 1 kHz. For self tracking, the Tx antennas 2,4,5,6 are used and appropriate antenna pairs are selected. For SAR imaging and 3D point cloud generation, the Tx antennas 1,2,3 and all 8 Rx antennas are used, resulting in an equivalent uniform linear array with 24 virtual elements. We implement all the MilliPoint design components in Matlab on the PC host, which uses the baseband signal samples as input and generates point clouds as output.

Setup. We co-locate a ZED stereo camera with the Radar to track its ground truth location as well as capture visual point clouds of the scenes in the experiment, as shown in Fig. 9. The Radar is placed vertically to achieve higher vertical accuracy for height estimation. To understand the performance of MilliPoint, we conduct controlled micro-benchmarks in an indoor spacious hall. The Radar is mounted on a cart to emulate a movable vehicle. Besides, case studies are conducted in an outdoor parking lot to demonstrate MilliPoint in practice. The Radar is mounted on the right front door of a car and we drive the car to pass by the target scenes. With both micro-benchmark and case study, we will show that MilliPoint improves the point cloud density and resolution

substantially compared with conventional solutions.



(a) (b)

Fig. 10. Performance of self tracking. (a) Tracking accuracy in different environments. (b) Tracking accuracy with different angular deviation of motion.

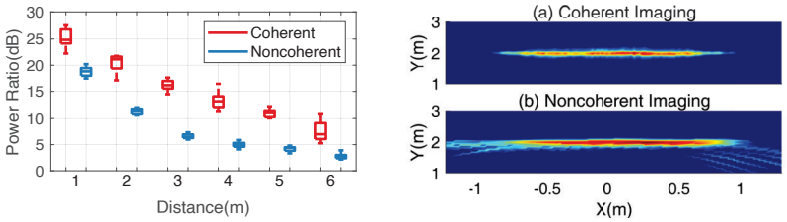
5.2 Micro-benchmarks

Self tracking accuracy. To measure the self tracking accuracy of MilliPoint, we mount the Radar on a cart and move it through a fixed distance of 5 m. The cart is guided by colored tape pasted on the ground. Due to lack of synchronization between the cart movement and the Radar, it is difficult to measure the instantaneous locations of the Radar where FMCW pulses are transmitted. We thus derive the overall moving distance and calculate the relative *cumulative* error, as the ratio between the absolute error and the ground truth moving distance. As comparison, we also measure the tracking accuracy of the ZED camera. We conduct experiments in various indoor and outdoor environment with different number of objects in front of the devices, including (1) a lab furnished with desks and equipment, (2) a hall with open space, (3) a parking lot with multiple cars and motorcycles, and (4) a road segment with sparse trees and street lamps on the side. The lab and the parking lot represent the cases with richful objects while the hall and the road represent the cases with few objects.

Fig. 10a shows the tracking accuracy in box plot. The group label "in" and "out" represent indoor and outdoor environment respectively; "w/" indicates that richful prominent reflecting objects exist in the scenes, while "wo" indicates very few objects besides walls and the ground. We observe that the median cumulative tracking error of MilliPoint is 1.2%–only 0.2% higher than that of the ZED camera which has a precision rating of sub mm. Besides, the two sensor modalities perform differently in different environment. MilliPoint has relatively larger tracking error in an outdoor environment where fewer and weaker reflecting objects exist and the channel correlation profile is noisier, leading to erroneous speed estimation. In contrast, camera has larger tracking errors when fewer representative anchors (*e.g.*, corners) appear in its FoV. For example, the wall in the hall and the close cars in the parking lot are too homogeneous to provide sufficient anchors for accurate location change estimation. In practice, the two sensor modalities can be combined to avoid large tracking errors.

Due to minor installation error, the antenna array of the vehicle Radar may not be strictly in parallel with the moving direction of the vehicle. To evaluate the impact of this angular deviation, we measure the tracking accuracy of MilliPoint while rotating the Radar by different tilting angles. The experiment is conducted in the lab. As shown in Fig. 10b, the relative tracking error increases with the increase of the angular deviation. Fortunately, the tracking error remains below 5% when the deviated angle is smaller than 8°, within which can the deviation be controlled even by manual calibration of drivers.

Target imaging quality. We proceed to evaluate the imaging performance of MilliPoint. To better control variables that impact the imaging process, we use flat metal boards that are 1.2 m wide and 2 m high as target objects. A non-coherent imaging approach is implemented for comparison. Specifically, we divide the cross-range aperture into segments of 20 cm length. For each segment, we generate one image where each pixel is assigned by the correlation between the steering vector derived from the relative location between the pixel and the midpoint

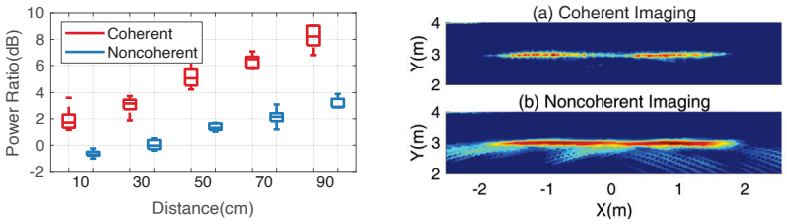


(a) (b)

Fig. 11. Performance of imaging accuracy. (a) Imaging SNR of objects at different distances. (b) An example of imaging result.

of this segment and the array response at the range of the pixel. Finally, images of all segments are max pooled to yield the final image. Although the non-coherent approach uses digital beamforming, it has effectively the same imaging resolution as the state-of-the-art non-coherent image approaches with phased array beamforming [69]. To evaluate the imaging quality, we generate a mask that is the ground truth location of the board, and define the ***power ratio*** between pixels within the mask and those out of the mask. Intuitively, if the object is correctly localized and its dimensions are correctly imaged, the power ratio is likely to be high.

First, we place one metal board in front of the Radar and in parallel with the cross-range aperture to evaluate the performance of imaging a single object. The distance between the object and the Radar aperture is varied from 1 m to 6 m. As shown in Fig. 11a, the power ratios of both imaging approaches decrease with the increase of the distance. The reasons are two folds: *(i)* The imaging resolution reduces, leading to a wider image than the real object. *(ii)* The reflected signal tends to be weaker, leading to the decrease of the power ratio. However, *the power ratio of MilliPoint is consistently higher than that of the non-coherent imaging, meaning that MilliPoint yields higher imaging quality.* Fig. 11b shows examples of imaging a board 2 m away. The imaging result of MilliPoint aligns with the location and width of the real object, with no side lobes or ghost images.

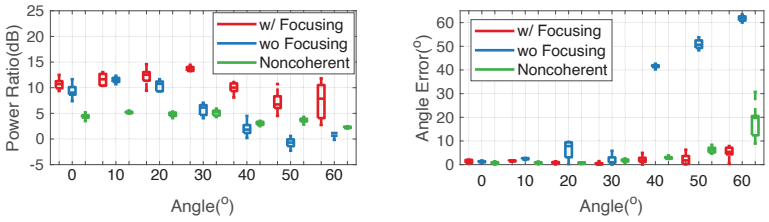


(a) (b)

Fig. 12. Performance of imaging resolution. (a) Imaging SNR of two objects with different spacing. (b) An example of imaging result.

Second, we place two metal boards side by side and 3 m away from the Radar aperture to further evaluate the *imaging resolution*. The space between the two boards is varied from 10 cm to 90 cm. As shown in Fig. 12a, the power ratios of both approaches gradually increase as the two boards separate further. The main reason is that the two boards with larger spacing can be more clearly differentiated in the image, and the mutual leakage decreases. *In comparison to the non-coherent approach, MilliPoint consistently leads to higher power ratio, i.e., sharper image and higher resolution.* Note that even when the space is only 10 cm, the power ratio of MilliPoint is positive at 2 dB, while that of the non-coherent approach is below 0, meaning that two objects' images are blurred into one, making any post processing algorithms impossible (*e.g.*, object segmentation). Fig. 12b shows an example of imaging two boards whose space is 50 cm. The two boards can be clearly segmented in the imaging

result of MilliPoint, whereas the non-coherent imaging leads to indistinguishable boundaries and large sections of ghost pixels due to its low angular resolution.

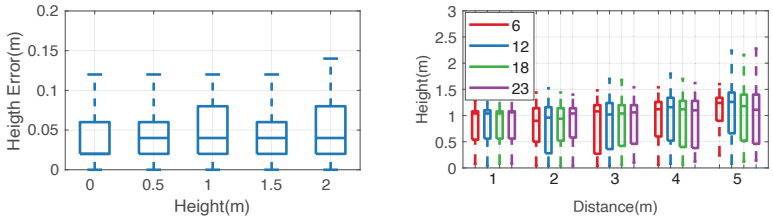


(a) (b)

Fig. 13. Performance of target detection and focusing. (a) Imaging SNR of objects with different orientations. (b) Accuracy of object orientations.

Effectiveness of target focusing. We now evaluate the necessity and effectiveness of MilliPoint's adaptive focusing mechanism. We place a board 3 m away from the cross-range aperture and vary their incident angle from 0 to 60°. Fig. 13a shows the power ratio of imaging results from MilliPoint with and without focusing, in comparison with the non-coherent approach. While the power ratios of the three decrease at larger incident angles, their decreasing trends are different. The degradation of the non-coherent approach is the slightest, and that of the original MilliPoint is more significant when the incident angle is large, due to the shortening of the equivalent aperture in parallel with the board. In contrast, the performance of MilliPoint without focusing degrades dramatically when the incidental angle exceeds 20°, since only the edge of the board can be correctly imaged due to specular reflection.

To further demonstrate the failure without focusing, we calculate orientations of the object's images and compare them with the ground truth. Specifically, we select prominent pixels, weight them with their values, and fit them with a line. We then derive the incidental of the line relative to the Radar aperture. As shown in Fig. 13b, without focusing, MilliPoint totally fails to image the board when the incident angle is over 40°, where the absolute errors are close to the ground truth of incident angles, meaning that only the parts in parallel with the cross-range direction can be imaged. In contrast, MilliPoint achieves consistently high accuracy in identifying the object's orientation.



(a) (b)

Fig. 14. Performance of height estimation for 3D point cloud generation. (a) Accuracy of height estimation. (b) Sensing range of height at different distances.

Accuracy of height estimation. To evaluate the estimation accuracy of point height, we use a hollow metal cylinder as the target, whose length is 1 m and diameter 15 cm. We lift the object by fixing two height-adjustable tripods on its both ends. The relative height between the object and the Radar is varied from 0 to 2 m, and the horizontal distance is fixed to 2 m. After generating point clouds, we filter out points of the static background, including the walls, ground, ceiling, and tripods, and calculate the average height of the remaining points

for evaluation. As shown in Fig. 14a, the median height errors are below 5 cm for all relative height settings, demonstrating the high accuracy of height estimation of MilliPoint.

Our Radar comprises patch antenna elements which themselves are directional and have limited field-of-view (FoV, approximately ±30° [24]). the point cloud generated by MilliPoint may not fully cover the target object. To evaluate the coverage of point clouds over objects along the height dimension, we let the Radar face towards a high wall and vary the distance from 1 m to 5 m. To show the impact of array size, we vary the number of antennas from 6 to 23. Upon generating the point clouds of the wall, we remove occasional outliers that are above the ceiling or below the ground. From the results in Fig. 14b, we have two observations. *First*, the height range is significantly smaller at shorter distances (*e.g.*, 1 m), since the Radar's FoV can only illuminate part of the wall. When the distance increases, the height range gradually approaches the actual height of the wall. *Second*, the height coverage (i.e., length of boxes in Fig. 14b) increases with the number of antennas used, as more prominent scatters at different heights can be separated as the resolution improves.

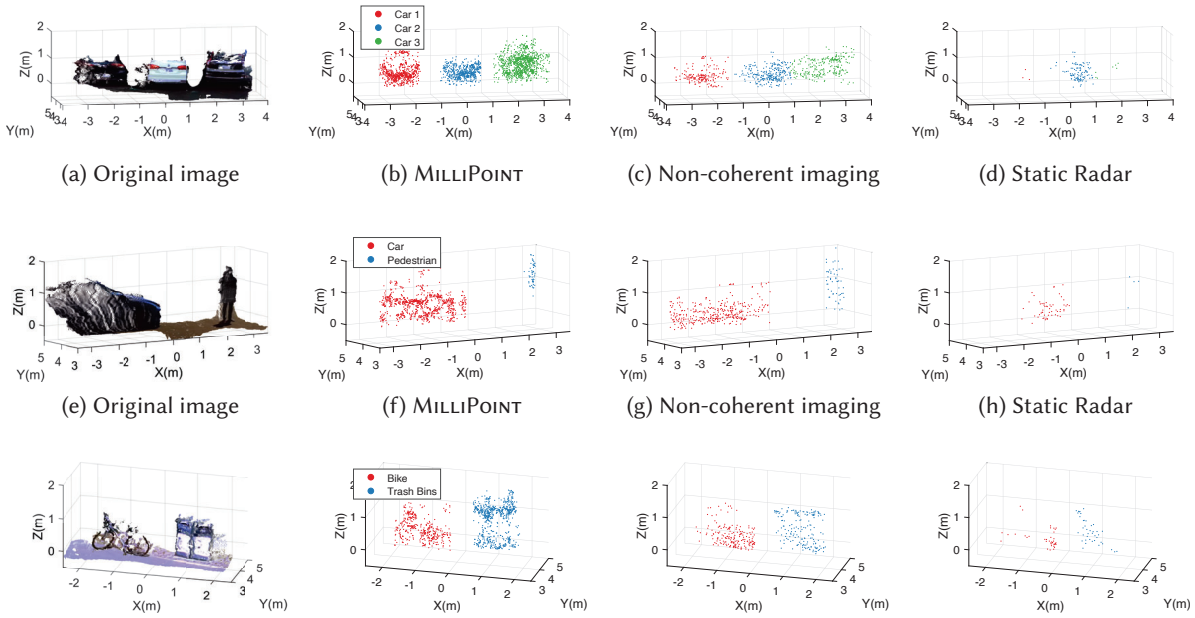
5.3 Case Study

To demonstrate the quality of the point clouds generated by MilliPoint, we conduct experiments in 3 representative outdoor scenarios (Fig. 15) with typical objects on roads, *e.g.*, cars, pedestrians, bikes, and trash bins. We mount the Radar on a car and drive the car along a cross-range aperture large enough to cover the objects of interest in each scene. The ZED camera is co-located with the Radar to capture visual ground truth. We compare MilliPoint with two alternative solutions used by the TI Radar: *(i) Static Radar:* a standalone radar that generates points when it is located in the middle of the cross-range aperture. *(ii) Non-coherent imaging:* combining the points non-coherently across locations, similar to [69]. Both approaches generate points by identifying the distance and 2D angles of scatters. More specifically, upon receiving signals, the Radar runs 1D FFT over time domain and applies the CFAR algorithm [46] to detect the distances where prominent scatters are. Then, for each distance value, the Radar combines the corresponding samples from all antennas and applies the Capon algorithm to obtain the 2D AoA (*i.e.*, azimuth and elevation) of the scatter. On the TI Radar, (Fig. 9), all 6 Tx and 8 Rx antennas equivalently constitute a 2D array with 48 elements, which can estimate 2D AoA. Finally, with both distance and AoA information, the Radar derives the Cartesian coordinate of the scatters.

Fig. 15 shows the resulting point clouds. We use the point cloud function of the ZED camera to generate visual point clouds from the stereo images. Due to strong ambient sunlight and fewer feature points on homogeneous surfaces, the points generated by the ZED camera may have large location errors. Thus, we select the view where the projection has the least distortion for each case to better demonstrate the results. We see that *MilliPoint generates the densest point clouds in all scenarios. The points are accurately aligned with the objects and reflect their shapes.* For example, in the first case, the trunks of the three cars are imaged and their relative sizes match the ground truth. In the second case, the whole side of the car is captured. In the last case, the mainframe and crooked front wheel of the bike are accurately shown. In contrast, the non-coherent point clouds are sparser and concentrate on the Radar's horizontal plane, due to the limited resolution of the small antenna array. Further, due to limited FoV, a static Radar can only generate several points for each object, which barely provides any information about the scene.

Despite its ability to achieve high resolution and overcome the specular reflection effect along the horizontal direction, MilliPoint still faces the problems of low resolution and specular reflection along the vertical direction, due to the limitation of the physical array aperture size on the Radar. Thus, some parts of the objects may be missed by MilliPoint. For example, the legs of the pedestrian in the second case are not imaged, since they specularly reflect signals to other directions [1]. Nonetheless, MilliPoint already captures representative partial shapes of major objects on the road, making them distinguishable.

Unlike optical sensors, the Radar-generated point cloud contains not only the location but also reflection

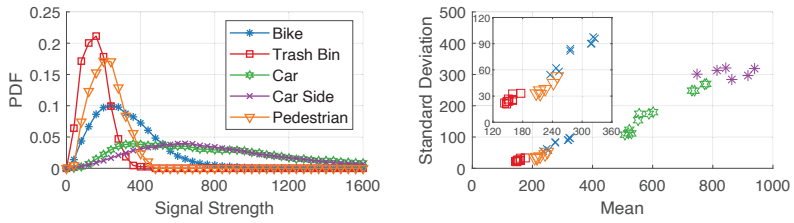


(a) Original image (b) MilliPoint (c) Non-coherent imaging (d) Static Radar

(e) Original image (f) MilliPoint (g) Non-coherent imaging (h) Static Radar

(i) Original image (j) MilliPoint (k) Non-coherent imaging (l) Static Radar

Fig. 15. Field test of point cloud generation. (a)-(d): A row of cars. (e)-(h): The side of a car and a pedestrian. (i)-(l): The side of a road with bike, tree and trash bins.



(a) (b)

Fig. 16. Strength of reflecting points on different targets. (a) Strength distribution. (b) Mean and standard deviation.

strength of each scatter point which indicates the target's material type. We verify this capability by plotting the signal strength distributions of the point clouds. As shown in Fig. 16a, the distribution varies a lot across objects. As most distributions are approximately Gaussian, we further calculate the mean and *std.* of the distributions. As shown in Fig. 16b, the same type of objects are closely clustered in the mean-std. feature domain. Cars have the largest feature values. Other objects are less separable but still exhibit a strong clustering effect. In summary, *the point clouds generated by MilliPoint embody not only the shape, but also the material information, and thus can be further exploited for object recognition.* More advanced post-processing mechanisms are left for our future work.

6 DISCUSSION

Quasistatic requirement of SAR. MilliPoint exploits SAR imaging to improve the Radar resolution and generates denser point clouds. Thus, the scene has to be quasistatic when the Radar moves along the cross-range

direction and generates SAR images. If objects in the scene move at speeds comparable to the Radar, their images will distort and virtual images will appear. However, MilliPoint is designed for surround views, where objects with similar speeds as the Radar are less common. Moreover, the Radar can detect moving objects by monitoring their relative Doppler frequency shifts and compensate their motions accordingly [40, 60]. We leave the integration of the motion compensation algorithm for moving objects as future work.

Potential applications. MilliPoint is designed as a generous approach to improve the resolution of mmWave Radar. Besides autonomous driving, MilliPoint can be applied to various applications. For example, a robot equipped with Radar can cruise in a building and reconstruct fine-grained 3D point clouds of indoor environments with MilliPoint. Besides, during fire disasters where light sensors such as camera and Lidar fail, rescue robots can exploit MilliPoint to navigate in disaster scenes, search, rescue, and evacuate victims. Last but not least, MilliPoint can be potentially used for the security check to detect concealed objects.

Real-time processing. We implement MilliPoint in Matlab and mainly demonstrate its feasibility and effectiveness. The current implementation cannot process Radar data in real-time. Specifically, on average, it takes about 0.5, 0.61, and 0.47 s to complete self tracking, automatic multi-focusing, and 3D point cloud generation respectively for 1 s Radar data. Nonetheless, MilliPoint can be reimplemented with more efficient languages such as C++/C and further optimized to run in real-time. For example, multiple steps can be parallelized, e.g., calculation of correlation profiles of multiple antenna pairs in self tracking, SAR imaging of multiple objects in automatic multi-focusing, and height estimation of points in 3D point cloud generation.

7 RELATED WORK

Automotive sensing technologies. Existing automotive sensing technologies can be divided into three categories based on the sensor type. First, a Lidar detects objects' locations by transmitting laser pulses and measuring the ToF and angles. By vertically stacking multiple laser channels, and horizontally sweeping narrow laser beams, Lidar can generate fine-grained 3D point clouds [19]. Cost and weather sensitivity aside, Lidar's point clouds have many unique properties compared with conventional 2D images, such as sparsity [67] and disorderliness [41]. Existing work extracted features from Lidar point cloud, either through hand-crafted feature engineering [9, 26, 52, 56] or automated deep learning [11, 32, 33, 42, 54], for semantic points segmentation and 3D object recognition. These approaches can potentially be adapted to the point clouds generated by MilliPoint.

Second, a camera retrieves color and even depth of high-resolution pixels, and understand objects better in comparison with other sensors. A vast literature exists in camera-based automobile perception algorithms, including object detection [12, 18, 23, 45], pose estimation [8, 22, 35, 55], and self-navigation [3, 36]. Camera images are also amenable for recognizing traffic signs [20], lights [28] and lanes [49], which are critical for self-driving. However, cameras are sensitive to lighting variation and cannot function in dark environments, which prevents them from being a standalone sensor modality for autonomous driving.

Third, Facilitated by antenna array processing, Radar can obtain the angular information of reflecting points which, combined with ToF, can localize the points in 3D space. Radar signals are less affected by weather conditions and insensitive to lighting [17]. However, due to the form factor constraints of the antenna array, conventional vehicle Radar suffers from low resolution [51] and specular reflection [13], and can only generate sparse and partial point clouds of objects, as demonstrated in our field tests (Sec. 5.3). RSA [70] and Ulysses [69] move the Radar and combines signal parameters (*e.g.*, AoA, AoD, and RSS) of specular reflections along the traces to reconstruct 2D object surfaces. However, due to aperture motion error, they only incoherently combine RF readings and cannot fundamentally improve imaging resolution. In contrast, MilliPoint develops a self-tracking algorithm with millimeter precision and enables SAR imaging [6] with finer resolution and generates 3D point clouds for autonomous driving applications. A recent system HawkEye [21] uses cGAN to generate high resolution depth image from low resolution and sparse Radar heatmaps for cars. However, it tends to overfit a single object

type since the training is done exclusively on a vehicle dataset. In contrast, MilliPoint is a closed-form approach designed for general objects on road.

Wireless ubiquitous sensing with potential vehicular use cases. Car-oriented ubiquitous sensing schemes have been widely explored [37, 48, 64]. FarSight [37] uses images taken by a smartphone to detect and track front cars for safety driving. In [48], a data-driven misbehavior detection system is proposed to predict driving states of vehicles and detect false information shared by malicious vehicles in a vehicular network. CoSense [64] learns the mobility patterns of personal vehicles and crowdsources with connected commercial vehicles to infer locations of personal vehicles at urban scale in real time.

On the other hand, wireless technologies, such as Wi-Fi [30, 31, 63, 66] and WiGig [16, 47, 58, 59], emerge as novel ubiquitous sensing modality. For example, FarSense [66] exploits the relative signal phase of two antennas to robustly monitor human respiration. Indotrack [31] enables passive human tracking by estimating AoA and Doppler frequency shift of the Wi-Fi signals reflected by the target. Due to the long wavelength and limited antenna aperture, these systems are unable to provide sufficient resolution for object sensing. Using 60 GHz WiGig radios, mTrack [58] can track object movement with millimeter precision. EMi [59] further takes advantage of the multipath reflections to reconstruct a coarse outline of an environment through a sparse set of sampling locations. However, these systems still fall short of resolution as they only pinpoint the dominant reflections in the environment.

8 CONCLUSION

SAR imaging has been adopted in airborne, orbital, and security applications. However, it faces fundamental challenges when applied to the consumer-grade vehicle Radar. Our MilliPoint system marks the first step in identifying such problems and developing solutions that are provable in realistic real-world 3D scenes. By enabling all-weather 3D point cloud generation, MilliPoint can become an alternative or complementary solution to the costly Lidar devices. The accumulation of Radar imaging data may also inspire new perception algorithms that cross the boundary of RF sensing and machine vision.

ACKNOWLEDGMENTS

We sincerely thank the anonymous reviewers and editors for their insightful feedback. This work was supported in part by the US National Science Foundation through NSF CNS-1854472, CNS-1901048, CNS-1952942, CNS-1925767.

REFERENCES

[1] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Fredo Durand. 2015. RF-Capture: Capturing the Human Figure Through a Wall. *ACM SIGGRAPH Asia* (2015).

[2] T. E. Bishop and P. Favaro. 2012. The Light Field Camera: Extended Depth of Field, Aliasing, and Superresolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 5 (2012).

[3] Sean L Bowman, Nikolay Atanasov, Kostas Daniilidis, and George J Pappas. 2017. Probabilistic data association for semantic slam. In *Proceedings of IEEE ICRA*.

[4] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. 2020. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of IEEE CVPR*.

[5] Jack Capon. 1969. High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE* 57, 8 (1969), 1408–1418.

[6] W Carrara, R Goodman, and R Majewski. 1995. *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms*. Norwood, MA, USA: Artech House.

[7] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. 2010. An introduction to total variation for image analysis. *Theoretical foundations and numerical methods for sparse recovery* 9, 263-340 (2010), 227.

[8] Xiaozhi Chen, Kaustav Kundu, Ziyu Zhang, Huimin Ma, Sanja Fidler, and Raquel Urtasun. 2016. Monocular 3d object detection for autonomous driving. In *Proceedings of IEEE CVPR*.

[9] Changhyun Choi, Yuichi Taguchi, Oncel Tuzel, Ming-Yu Liu, and Srikumar Ramalingam. 2012. Voting-based pose estimation for robotic

assembly using a 3D sensor. In *Proceedings of IEEE ICRA*.
[10] J. Dickmann, N. Appenrodt, J. Klappstein, H. Bloecher, M. Muntzinger, A. Sailer, M. Hahn, and C. Brenk. 2015. Making Bertha See Even More: Radar Contribution. *IEEE Access* (2015).
[11] Yi Fang, Jin Xie, Guoxian Dai, Meng Wang, Fan Zhu, Tiantian Xu, and Edward Wong. 2015. 3d deep shape descriptor. In *Proceedings of IEEE CVPR*.
[12] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. 2009. Object detection with discriminatively trained part-based models. *IEEE TPAMI* 32, 9 (2009), 1627–1645.
[13] Alex Foessel, Sachin Chheda, and Dimitrios Apostolopoulos. 1999. Short-range millimeter-wave radar perception in a polar environment. (1999).
[14] Rocco Furferi, Lapo Governi, Yary Volpe, and Monica Carfagni. 2013. Design and assessment of a machine vision system for automatic vehicle wheel alignment. *International Journal of Advanced Robotic Systems* 10, 5 (2013), 242.
[15] Paul Gao, Hans-Werner Kaas, Det Mohr, and Dominik Wee. 2016. Automotive revolution–perspective towards 2030: How the convergence of disruptive technology-driven trends could transform the auto industry. *Advanced Industries, McKinsey & Company* (2016).
[16] Dolores Garcia Marti, Jesús Omar Lacruz, Pablo Jimenez Mateo, and Joerg Widmer. 2020. POLAR: Passive object localization with IEEE 802.11 ad using phased antenna arrays. (2020).
[17] Axel Gern, Uwe Franke, and Paul Levi. 2001. Robust vehicle tracking fusing radar and vision. In *Proceedings of IEEE MFI*.
[18] Ross Girshick. 2015. Fast r-cnn. In *Proceedings of IEEE ICCV*.
[19] Craig Glennie and Derek D Lichti. 2010. Static calibration and analysis of the Velodyne HDL-64E S2 for high accuracy mobile scanning. *MDPI Remote Sensing* 2, 6 (2010), 1610–1624.
[20] Jack Greenhalgh and Majid Mirmehdi. 2014. Recognizing text-based traffic signs. *IEEE TITS* 16, 3 (2014), 1360–1369.
[21] Junfeng Guan, Sohrab Madani, Suraj Jog, Saurabh Gupta, and Haitham Hassanieh. 2020. Through Fog High-Resolution Imaging Using Millimeter Wave Radar. In *Proceedings of the IEEE CVPR*.
[22] Ishan Gupta, Akshay Rangesh, and Mohan Trivedi. 2018. 3D Bounding Boxes for Road Vehicles: A One-Stage, Localization Prioritized Approach using Single Monocular Images.. In *Proceedings of Springer ECCV*.
[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of IEEE CVPR*.
[24] Texas Instruments. 2019. AWR1642 obstacle detection sensor with wide field-of-view (FOV) antenna evaluation module. http://www.ti.com/tool/AWR1642BOOST-ODS
[25] SAE International. 2014. Automated driving: levels of driving automation are defined in new SAE international standard J3016.
[26] Andrew E. Johnson and Martial Hebert. 1999. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE TPAMI* 21, 5 (1999), 433–449.
[27] Jaime Laviada, Ana Arboleya-Arboleya, Yuri Álvarez, Borja González-Valdés, and Fernando Las-Heras. 2017. Multiview three-dimensional reconstruction by millimetre-wave portable camera. *Nature Scientific reports* 7, 1 (2017), 6479.
[28] Jesse Levinson, Jake Askeland, Jennifer Dolson, and Sebastian Thrun. 2011. Traffic light mapping, localization, and state detection for autonomous vehicles. In *Proceedings of IEEE ICRA*.
[29] Andreas Ley, Olivier D'Hondt, and Olaf Hellwich. 2018. Regularization and completion of TomoSAR point clouds in a projected height map domain. *IEEE JSTARS* 11, 6 (2018), 2104–2114.
[30] Hong Li, Wei Yang, Jianxin Wang, Yang Xu, and Liusheng Huang. 2016. WiFinger: talk to your smart devices with finger-grained gesture. In *Proceedings of ACM UbiComp*.
[31] Xiang Li, Daqing Zhang, Qin Lv, Jie Xiong, Shengjie Li, Yue Zhang, and Hong Mei. 2017. IndoTrack: Device-free indoor human tracking with commodity Wi-Fi. *Proceedings of ACM IMWUT* 1, 3 (2017), 1–22.
[32] Wenjie Luo, Bin Yang, and Raquel Urtasun. 2018. Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net. In *Proceedings of IEEE CVPR*.
[33] Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. 2015. Geodesic convolutional neural networks on riemannian manifolds. In *Proceedings of IEEE ICCV Workshops*.
[34] Theo Moons, Luc Van Gool, and Maarten Vergauwen. 2010. 3D Reconstruction from Multiple Images Part 1: Principles. *Foundations and Trends in Computer Graphics and Vision* 4, 4 (2010).
[35] Arsalan Mousavian, Dragomir Anguelov, John Flynn, and Jana Kosecka. 2017. 3d bounding box estimation using deep learning and geometry. In *Proceedings of IEEE CVPR*.
[36] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. 2015. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE T-RO* 31, 5 (2015), 1147–1163.
[37] Akshay Uttama Nambi, Adtiya Virmani, and Venkata N Padmanabhan. 2018. FarSight: A Smartphone-based Vehicle Ranging System. *Proceedings of ACM IMWUT* 2, 4 (2018), 1–22.
[38] Jeffrey A. Nanzer. 2013. Microwave and Millimeter-Wave Remote Sensing for Security Applications. (2013).
[39] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, Pat Hanrahan, et al. 2005. Light field photography with a hand-held

plenoptic camera. *Computer Science Technical Report CSTR* 2, 11 (2005), 1–11.
[40] RP Perry, RC Dipietro, and RL Fante. 1999. SAR imaging of moving targets. *IEEE Trans. Aerospace Electron. Systems* 35, 1 (1999), 188–200.
[41] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of IEEE CVPR*.
[42] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Proceedings of ACM NeurIPS*.
[43] Kun Qian, Chenshu Wu, Zheng Yang, Yunhao Liu, and Kyle Jamieson. 2017. Widar: Decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi. In *Proceedings of ACM MobiHoc*.
[44] NavTech Radar. 2019. ClearWay Software and Sensors. https://navtechradar.com/clearway-technical-specifications/
[45] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of IEEE CVPR*.
[46] Mark A Richards. 2005. *Fundamentals of radar signal processing*. Tata McGraw-Hill Education.
[47] Panneer Selvam Santhalingam, Al Amin Hosain, Ding Zhang, Parth Pathak, Huzefa Rangwala, and Raja Kushalnagar. 2020. mmASL: Environment-Independent ASL Gesture Recognition Using 60 GHz Millimeter-wave Signals. *Proceedings of ACM IMWUT* 4, 1 (2020), 1–30.
[48] Ankur Sarker and Haiying Shen. 2018. A Data-Driven Misbehavior Detection System for Connected Autonomous Vehicles. *Proceedings of ACM IMWUT* 2, 4 (2018), 1–21.
[49] David Schreiber, Bram Alefs, and Markus Clabian. 2005. Single camera lane detection and tracking. In *Proceedings of IEEE ITSC*.
[50] David M Sheen, Douglas L McMakin, and Thomas E Hall. 2001. Three-dimensional millimeter-wave imaging for concealed weapon detection. *IEEE Transactions on microwave theory and techniques* 49, 9 (2001), 1581–1592.
[51] Merrill I Skolnik. 1962. Introduction to radar. *Radar handbook* 2 (1962), 21.
[52] Fridtjof Stein and Gérard Medioni. 1992. Structural indexing: Efficient 3-D object recognition. *IEEE TPAMI* 2 (1992), 125–145.
[53] StereoLabs. 2019. Zed. https://www.stereolabs.com/zed/
[54] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. 2015. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings IEEE CVPR*.
[55] Shubham Tulsiani and Jitendra Malik. 2015. Viewpoints and keypoints. In *Proceedings of IEEE CVPR*.
[56] Oncel Tuzel, Ming-Yu Liu, Yuichi Taguchi, and Arvind Raghunathan. 2014. Learning to rank 3d features. In *Proceedings of Springer ECCV*.
[57] Waymo. 2019. Waymo. https://waymo.com
[58] Teng Wei and Xinyu Zhang. 2015. MTrack: High-Precision Passive Tracking Using Millimeter Wave Radios. In *Proceedings of the ACM Annual International Conference on Mobile Computing and Networking (MobiCom)*. Association for Computing Machinery.
[59] Teng Wei, Anfu Zhou, and Xinyu Zhang. 2017. Facilitating Robust 60 GHz Network Deployment by Sensing Ambient Reflectors. In *Proceedings of the 14th USENIX Conference on Networked Systems Design and Implementation (NSDI)*.
[60] SUSANAS Werness, WALTERG Carrara, LS Joyce, and DAVIDB Franczak. [n.d.]. Moving target imaging algorithm for SAR data. *IEEE Trans. Aerospace Electron. Systems* 26, 1 ([n. d.]), 57–67.
[61] Wikipedia. 2016. Tesla Autopilot. https://en.wikipedia.org/wiki/Tesla_Autopilot#Handan,_China_(January_20,_2016)
[62] Wikipedia. 2018. Death of Elaine Herzberg. https://en.wikipedia.org/wiki/Death_of_Elaine_Herzberg
[63] Chenshu Wu, Feng Zhang, Yusen Fan, and KJ Ray Liu. 2019. RF-based inertial measurement. In *Proceedings of ACM SIGCOMM*.
[64] Xiaoyang Xie, Yu Yang, Zhihan Fang, Guang Wang, Fan Zhang, Fan Zhang, Yunhuai Liu, and Desheng Zhang. 2018. coSense: Collaborative urban-scale vehicle sensing based on heterogeneous fleets. *Proceedings of ACM IMWUT* 2, 4 (2018), 1–25.
[65] Jie Xiong and Kyle Jamieson. 2013. Arraytrack: A fine-grained indoor location system. In *roceedings of USENIX NSDI*.
[66] Youwei Zeng, Dan Wu, Jie Xiong, Enze Yi, Ruiyang Gao, and Daqing Zhang. 2019. FarSense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas. *Proceedings of ACM IMWUT* 3, 3 (2019), 1–26.
[67] Yin Zhou and Oncel Tuzel. 2018. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of IEEE CVPR*.
[68] Xiao Xiang Zhu and Richard Bamler. 2011. Demonstration of super-resolution for tomographic SAR imaging in urban environment. *IEEE TGRS* 50, 8 (2011), 3150–3157.
[69] Yanzi Zhu, Yuanshun Yao, Ben Y Zhao, and Haitao Zheng. 2017. Object Recognition and Navigation Using a Single Networking Device. In *Proceedings of ACM MobiSys*.
[70] Yanzi Zhu, Yibo Zhu, Ben Y Zhao, and Haitao Zheng. 2015. Reusing 60ghz radios for mobile radar imaging. In *Proceedings of ACM MobiCom*.
[71] J. Ziegler, P. Bender, M. Schreiber, H. Lategahn, T. Strauss, C. Stiller, T. Dang, U. Franke, N. Appenrodt, C. G. Keller, E. Kaus, R. G. Herrtwich, C. Rabe, D. Pfeiffer, F. Lindner, F. Stein, F. Erbs, M. Enzweiler, C. Knoppel, J. Hipp, M. Haueis, M. Trepte, C. Brenk, A. Tamke, M. Ghanaat, M. Braun, A. Joos, H. Fritz, H. Mock, M. Hein, and E. Zeeb. 2014. Making Bertha Drive—An Autonomous Journey on a Historic Route. *IEEE Intelligent Transportation Systems Magazine* 6, 2 (2014).