PanoRadar

Presenter: Sai Jogannagari

Background

- Robotics and Autonomous systems rely on accurate sensing and imaging
- Traditionally, LiDAR is preferred for detection and mapping
- This paper aims to use RF signals enhanced by signal processing and and machine learning for detection and mapping due to promising resilience against certain environmental conditions:
  - Dust
  - Fog
  - Smoke
  - Poor lighting

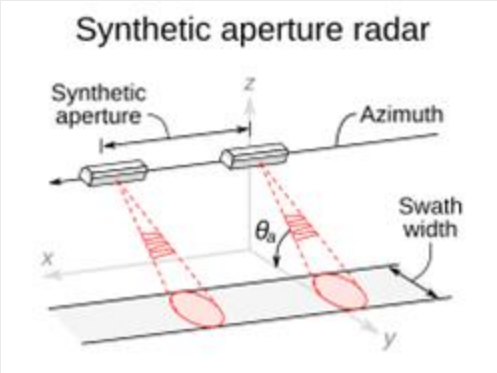
Challenges with using RF signals for sensing/imaging

- Poor resolution of RF sensors vs CMOS sensor (small array of sensors vs millions of pixels)
  - Limited angular resolution
- Azimuth/Cross-range resolution increases proportional to aperture of antenna (for a given wavelength) [1]

Related Work

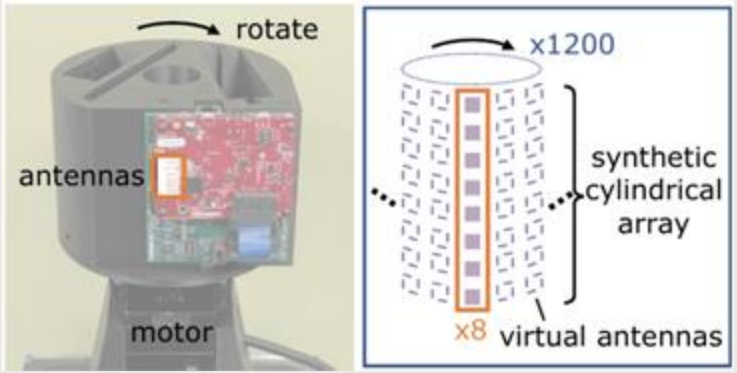
- Techniques for imaging using WiFI, RFID, radars, multi-sensor setups
- Synthetic Aperture Radar (SAR)
- Planar array emulation using horizontal and vertical sliders (limited by long scanning time)
- Circular SAR (mainly used in large static setups)
- 2D CNNs have been used but these do not develop 3D understanding of the surroundings
- 3D convolutions have been tried for 3D imaging but they are inefficient

These methods tried in the past are either not efficient enough or not well suited for mobile robotics applications



Central Design: Rotating Radar

- PanoRadar proposed solution: rotating vertical array of 8 sensors
  - Synthesizes 8x1200 virtual dense cylindrical array
- Benefits:
  - Expanded FOV
  - Low cost
  - Increased mobility
  - Fast capture time



Implementation Challenges: External Motion Tracking

- Combination of all the synthetic antenna requires sub wavelength precision location information
  - Not possible with IMUs or wheel odometers
- Leveraging RF signals for motion estimation is problematic due to ambiguity between Doppler effect and angle of arrival
- Proposed Solution: incorporate multiple reflectors and observe radial speed from different angles

Implementation Challenges: Limited Elevation Resolution

- Problem: There’s only 8 antennas along the vertical axis
- Claim: Indoor environments are more regular across the vertical dimension
- ML models can also analyze cross dimensional dependencies to interpolate in the vertical dimension and enhance the limited resolution

Implementation Challenges: 3d Learning

- 3D convolution is impractical as voxeling an (highly imbalanced) indoor space requires large amounts of memory and processing

Main Contributions

Main contribution is more efficient processing algorithm with cheaper/cots hardware while maintaining performance for mobile robotics use cases

1) 3d imaging with rotating radar
2) Motion estimation and compensation algorithm
3) Vertical and range resolution enhancement with ML
4) RF for various visual recognition tasks

Cylindrical Array Imaging

h^a = height of antenna a

r = radius of antennas

ω = angular speed of antennas

θ_d = azimuth of imaging direction d

ϕ_d = elevation of imaging direction d

S_a^t = intermediate frequency signals from antenna a at time t

- Range FFT to obtain range dimension
- Limit antenna positions used for

$$\vec{p}_t^a = (r\cos(\omega t), r\sin(\omega t), h^a). \quad (1)$$

Location of antenna a at time t

$$\vec{d} = (\cos\phi_d \cos\theta_d, \cos\phi_d \sin\theta_d, \sin\phi_d). \quad (2)$$

Desired imaging direction

$$B(\vec{d}) = \sum_{a,t} S_t^a \exp\left(j2\pi \frac{2\vec{d}\cdot\vec{p}_t^a}{\lambda}\right),$$

2D Beamformed Signal

Cylindrical Array Imaging Results

- Fine grained azimuth resolution (0.96 degrees for omnidirectional antennas)
- Fine grained range resolution (3.75 cm)
- Limited elevation resolution (14.2 degrees)

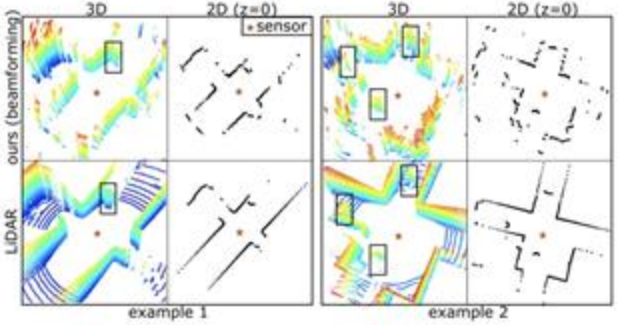


Figure 4: RF imaging results with a stationary robot. Our beamforming results capture humans in a rough shape, with limited elevation resolution.

Motion Estimation and Mapping

- Moving robot causes inaccuracies in beamforming due to RF image distortion
- Proposed Solution: Leverage Doppler effect in reflected signals to estimate robot motion
  - Complicated due to radar rotation and robot motion impacting AoA and Doppler effects

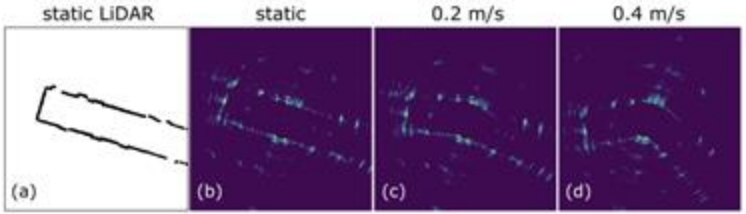


Figure 5: Distortion of the imaging results due to robot motion. 2D visualizations show that the distortion gets worse as the robot starts to move.

Decouple AoA and Doppler Effect

- Assume antenna rotates with radius r and angular velocity ω along z axis
- Assume antenna has linear velocity v (vector)
- Assume both velocities are constant over one rotation cycle (0.5s)

$$d(t) = R_n - \underbrace{r\cos(\omega t - \theta_n)}_{\text{radar rotation}} - \underbrace{vt\cos(\theta_v - \theta_n)}_{\text{robot motion}}, \quad (6)$$

Approximate distance between reflector and antenna at time t

$$d'(t, t_c) = d(t) + r\cos(\omega(t_c - t)). \quad (7)$$

Compensated distance at time t

$$d'(t, t_c) = r\omega t(\omega t_c - \theta_n) - vt\cos(\theta_v - \theta_n) + \text{const.}, \quad (8)$$

Approximate compensated distance at time t

Decouple AoA and Doppler Effect

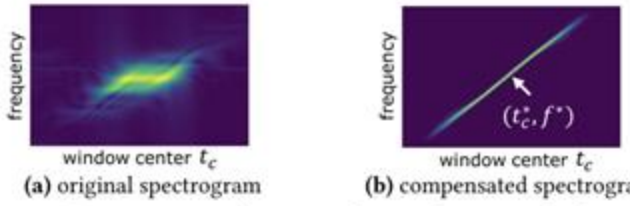
- Take slow time FFT of compensated antenna measurements to yield (9)

$$f(t_c) = \frac{2}{\lambda}[r\omega(\underbrace{\omega t_c - \theta_n}_{\text{AoA}}) - \underbrace{v\cos(\theta_v - \theta_n)}_{\text{Doppler speed}}], \quad (9)$$

- Strongest response appears when antenna is directly facing reflector

$$f^* = f(t_c^*) = -2v\cos(\theta_v - \omega t_c^*)/\lambda. \quad (10)$$

- ωt^*_c indicates direction of reflector
- $-\lambda f^*/2$ indicates Doppler speed



(a) original spectrogram **(b)** compensated spectrogram

Figure 6: RF spectrograms showing reflections from a single object. Each column of the spectrogram represents the slow-time FFT within a window centered at a specific t_c. The line in (b) follows Eqn. (9).

- Use Hough Transform with slope $2\omega^2 r/\lambda$ to identify lines in spectrogram

Robust Motion Estimation

- Measurement of AoA and Doppler speed from just a single reflector cannot accurately estimate v and θ_v
- Note that Doppler speed only reflects radial speed in direction of reflector and measurements can be very noisy
- Combining multiple reflectors by
  - Use range FFT to isolate reflectors at various distances
  - Multiple reflectors and lines are detected
  - All peaks fall on $f = -2v\cos(\theta_v - \omega t_c)/\lambda,$
  - Speed v and initial phase θ_v
- Perform sinusoidal curve fitting

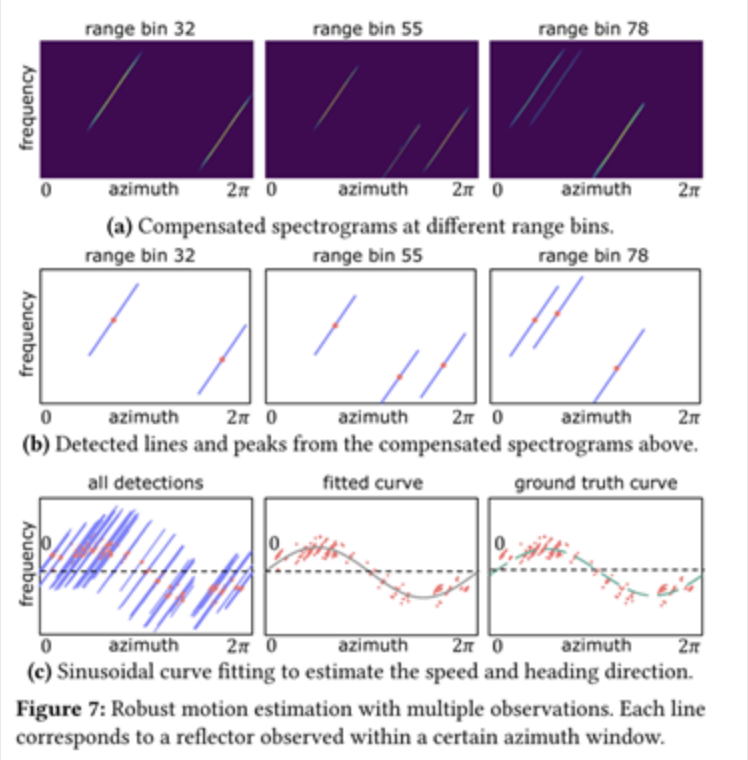


Figure 7: Robust motion estimation with multiple observations. Each line corresponds to a reflector observed within a certain azimuth window.

Efficient Compensation and Imaging

- Revise beamforming equation to account for robot motion in x, y plane

$$B(\vec{d},\vec{v}) = \sum_{a,t} S_t^a \exp\left(j2\pi \frac{2\vec{d}\cdot(\vec{p}_t^a + \vec{v}\cdot t)}{\lambda}\right). \quad (11)$$

- Complexity

Phase Steering. Using Delay-and-sum [30, 65] in our system would have a computational complexity of $O(\Theta\Phi WAN^2)$.

2D Beamforming. Using beamforming (11) followed by range FFT has a complexity of $O(\Theta\Phi WAN \log N)$.

Consecutive 1D Beamforming. We re-write (11) into:

$$B(\vec{d},\vec{v}) = \sum_t S'_t \exp\left(j4\pi\vec{d}'\cdot(\vec{p}'_t + \vec{v}'\cdot t)/\lambda\right),$$
$$S'_t = \sum_a S_t^a \exp\left(j4\pi h^a \sin\phi_d/\lambda\right), \quad (12)$$

where $\vec{d}'$, $\vec{p}'_t$ and $\vec{v}'$ are the first two dimensions of $\vec{d}$, $\vec{p}_t^a$ and $\vec{v}$, respectively. It shows that the compensation in elevation is independent of that in azimuth. This approach is effectively performing two consecutive steps of 1D beamforming, with a total complexity of $O(\Theta\Phi AN + \Theta\Phi WN \log N) = O(\Theta\Phi WN \log N)$ given that $W \gg A$.

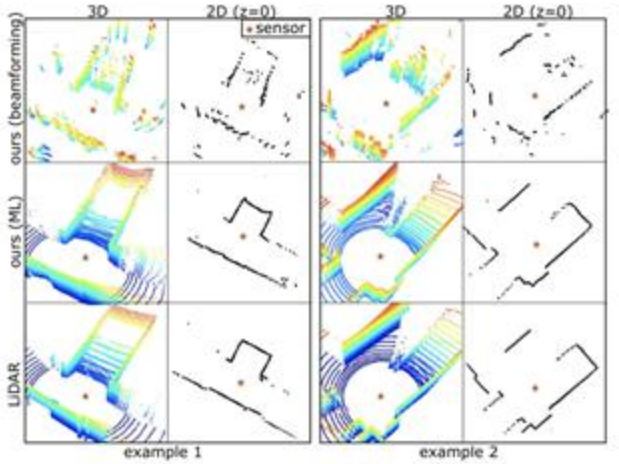


Figure 8: RF imaging results with a moving robot with beamforming (top) and with additional ML-based resolution enhancement (mid). Our motion estimation and compensation avoid the distortion, resulting in high range and azimuth resolutions. The 3D learning model further enhances the elevation resolution, showing detailed structures like stairs.

Enhanced Imaging with ML

- Cross dimensional info such as depth cues and gravity information potentially allow CNN resolution enhancement in the vertical axis

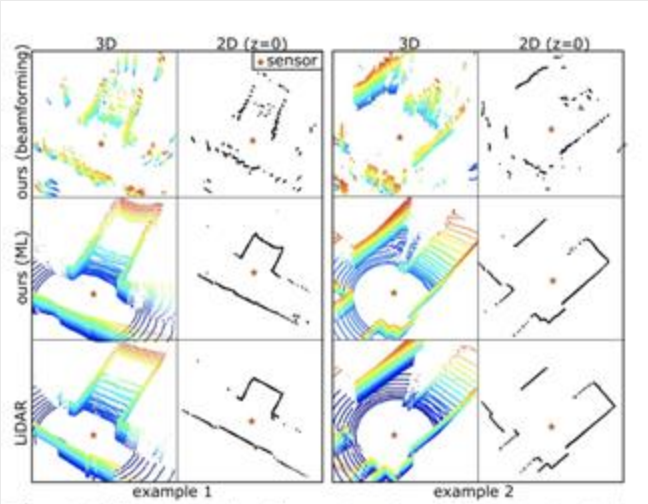


Figure 8: RF imaging results with a moving robot with beamforming (top) and with additional ML-based resolution enhancement (mid). Our motion estimation and compensation avoid the distortion, resulting in high range and azimuth resolutions. The 3D learning model further enhances the elevation resolution, showing detailed structures like stairs.

Resolution Enhancement with ML

- RF and LiDAR data used as inputs and targets for network training
- RF tensors are of size 512 x 64 x 256 (azimuth x elevation x range)
- LiDAR targets are 2D range maps of size 512 x 64
- Treat range as channel dimension and compress sparse signals across range dimension, enhancing efficiency and decreasing likelihood of overfitting
- Emulation of LiDAR decreases multipath effects (figure 10)
- Glass regions are masked to not misinform model by incorrect “ground truth” (figure 11)
- Use of perceptual loss in addition to L1 loss also retains higher frequency features (figure 12)

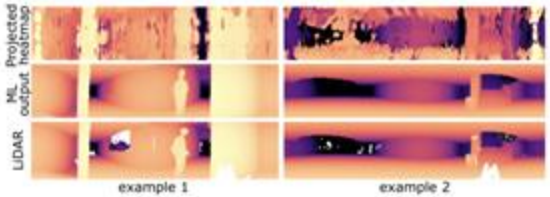


Figure 9: 3D RF heatmap (top) and the resolution enhancement model output (middle). To visualize a 3D heatmap, we take the range value of the peak in each direction and visualize the 3D heatmap as a 2D image.

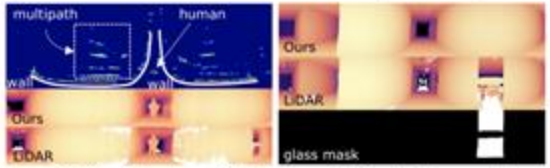


Figure 10: PanoRadar learns the first reflection under the presence of multipath interference.

Figure 11: PanoRadar properly handles glass regions (bottom, white pixels) and recovers their depth (top) while LiDAR fails to do so (middle).

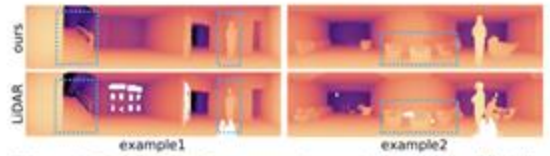


Figure 12: PanoRadar produces range images with fine-grained details.

Visual Recognition with ML

- An extra convolutional layer is used to predict surface normal vectors (used in visual perception tasks like SLAM), with LiDAR used to derive ground truth for training
- Use of pretrained ResNet-101 to predict 11 semantic classes
- Use of ResNet-101 with Feature Pyramid Network and Faster R-CNN for object detection
- Human Localization can be obtained as a byproduct of object detection

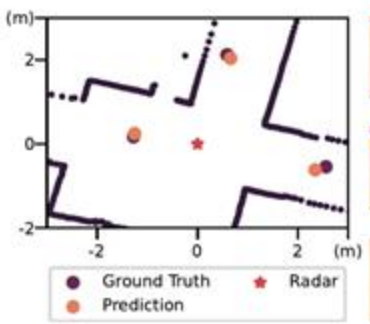


Figure 13: PanoRadar localizes humans on a 2D floor plan.

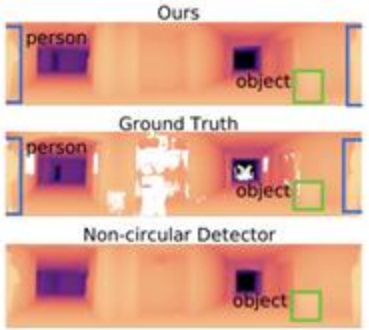


Figure 14: PanoRadar detects objects across boundary.

Panoramic Learning & Ground Truth Labels

- Input images are inherently panoramic so utilizing features that cross left and right boundaries can potentially improve model
- Panoramic learning can be leveraged by
  - Circular padding across azimuth
  - Disable bounding box clamping across azimuth
  - Allow IoU calculation to account for cross-boundary bounding boxes
  - Modify ROI pooling to first duplicate feature maps horizontally
- Ground Truth labels generated in the process of iteratively training a model and incorporating manual correction

Implementation & Dataset

- TI AWR1843 single-chip mmWave FMCW
- 4 GHz bandwidth sweep from 77 to 81 GHz
- Jetson Nano records raw samples
- Resolution enhancement model is structured into 7 stages each with 4 ResNet blocks
- Dataset spans 12 buildings constructed over the span of a century
- 11,033 synchronized RF and LiDAR data (total size 461GB after processing)

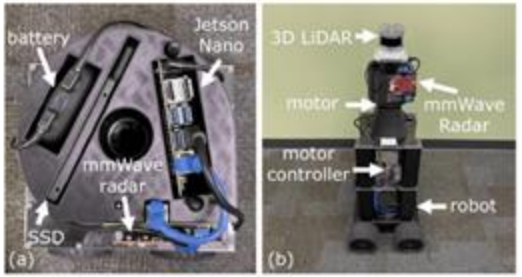


Figure 15: Implementation. Left: hardware setup that captures RF signals. Right: PanoRadar and LiDAR mounted on a mobile robot platform.

Evaluation

- Compared against LiDAR ground truth, model achieves mean absolute error of 15.76 cm
  - Point cloud (recovered from RF heatmap) accuracy is noticeably improved with ML enhancement
- Surface normal estimation: MAE of 8.83 degrees and median error of 2.17 degrees
- Semantic recognition: mIoU of 48.00
- Object detection: AP30 of 52.33 and AP50 of 38.30 (state of the art AP50 is 60+)
- Human Localization: Average error of 12.24 cm along range and 1.47 degrees along azimuth
- Qualitative performance: high quality generated images without some of the poor LiDAR artifacts

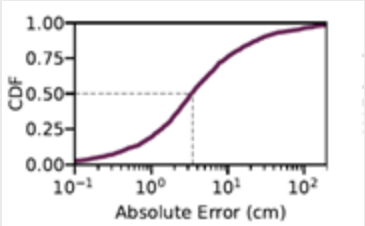


Figure 18: The CDF for absolute error of range image estimation.

Table 3: RF imaging point cloud error with and without machine learning. (CD: Chamfer Distance, HD: Modified Housdorff Distance)

	CD (2D)	CD (3D)	HD (2D)	HD (3D)
Beamforming only	21.4 cm	26.6 cm	12.8 cm	12.0 cm
Beamforming + ML	7.43 cm	6.96 cm	3.12 cm	3.23 cm

Evaluation

- Motion Estimation Accuracy (max speed 0.6m/s and avg speed 0.39m/s): 8.48mm/s speed estimation error and 1.09 direction estimation error
- Imaging Performance relatively robust to motion estimation
- Imaging performance drops at farther distances
- Circular model maintains performance against all orientations of image and object detection performance when bounding box is split across boundary

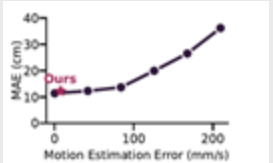


Figure 19: The effect of motion errors to imaging performance.

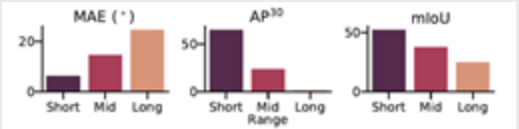


Figure 21: Performance of visual recognition at different distances: short (0-3m), mid (3-6m) and long (6-10m).

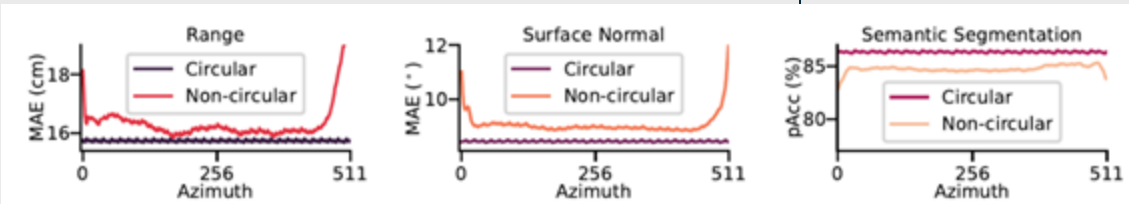


Figure 22: The column-wise metrics of circular and non-circular model across 512 azimuth columns for range estimation, surface normal estimation, and semantic segmentation.

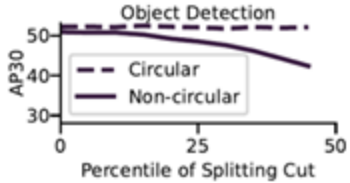


Figure 23: The AP^{30} verses split percentile in panoramic-rotate test.

Limitations, and Future Work

- Extend PanoRadar to other settings such as warehouses, malls, driving
- Repeat approach with better radar and achieve similar resolution with smaller ML model
- Work on a method to take advantage of multipath reflections while learning efficiently

“Opinion”

- Generally promising results - some of the results are not state of the art
- Doesn’t seem like there is anything revolutionary here

Summary of Perusall

https://app.perusall.com/courses/cos597e_f2025-advanced-topics-in-computer-science-neural-sensing-modeling-and-understanding/panoradar