

## 1. Motivation

### Value proposition

Robust 3D perception is essential for autonomous robots, AR/VR headsets, drones, and assistive devices.

- **Cameras & LiDAR** both fail in smoke, fog, snow, and dim light. LiDAR adds further cost, bulk, and power draw.
- **Specialized radar** can see through these conditions, but demands multi-GHz licensed spectrum and bespoke hardware that keep it out of compact platforms.
- **Rascene** reuses mmWave OFDM communication signals from 5G/Wi-Fi devices for scalable 3D imaging.

Tech.	Medium	Waveform	O.P.	O.R.	S.L.E.	Hardware	P.C.	Cost	Scalable
Camera	Light	--	No	Poor	--	Dedicated	Low	Low	High
LiDAR	Laser	Pulsed	No	Poor	--	Dedicated	High	High	Low
Radar	Radio	FMCW, etc.	Yes	Good	No	Dedicated	Med	Med	Low
Rascene	Radio	OFDM	Yes	Good	Yes	Reused	Low	~Zero	High

O.P.= Obstacle Penetration, O.R.= Occlusion Resilience, S.L.E.=Spectrum License Exempt, P.C.=Power Consumption

Key idea: turn the radios already in your phone, laptop, and access point into a 3D scene camera.  
 No new hardware, no new spectrum.

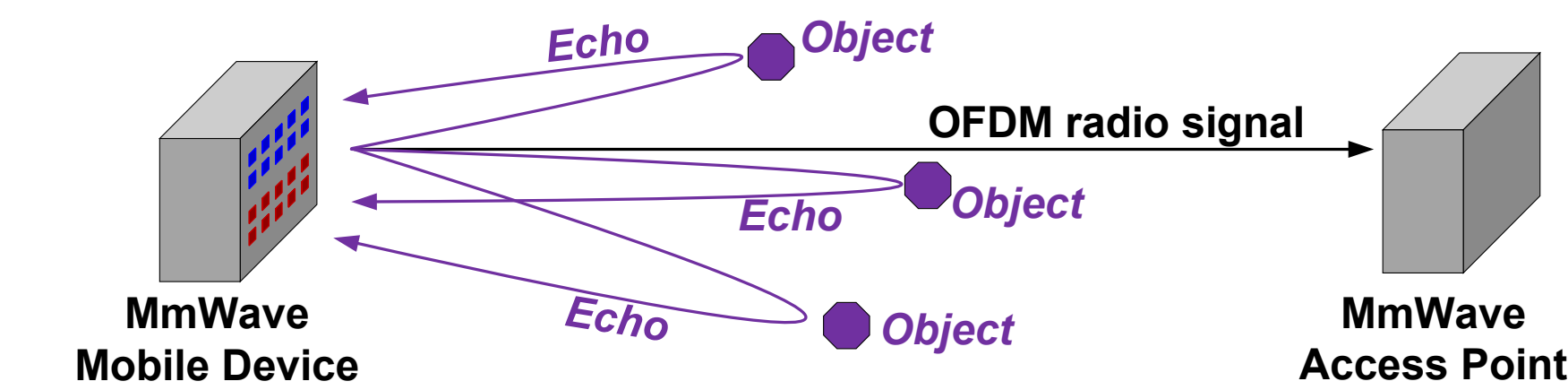
### Contributions

- **Monostatic ISAC framework for 3D imaging**  
 Rascene reconstructs 3D scenes from mmWave OFDM communication echoes on a single commodity device, with no dedicated radar hardware and no licensed spectrum.
- **Confidence-aware multi-frame RF fusion**  
 A source-driven projection fuses arbitrarily posed RF frames, reinforcing true geometry while suppressing multipath ghosts.
- **Real-world RF and LiDAR dataset**  
 A synchronized dataset across 20 indoor environments shows strong cross-scene generalization, robustness to occluders, and state-of-the-art within-dataset accuracy.

## 2. Method Overview

### A. Hardware: OFDM echoes to radio point clouds

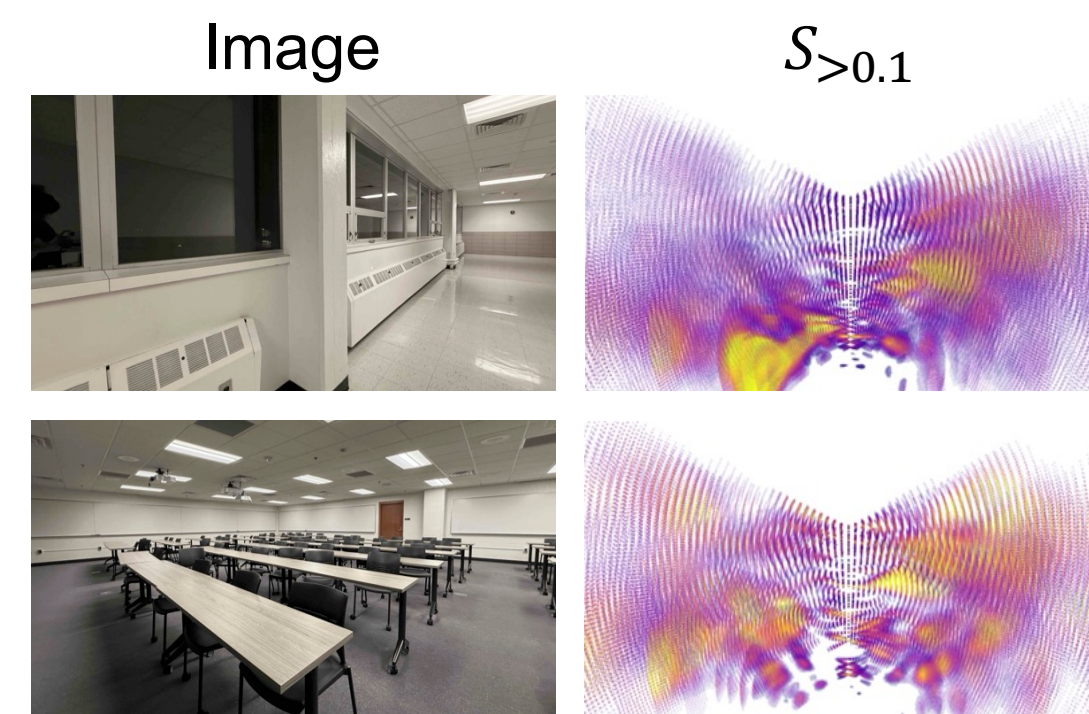
#### Monostatic sensing setup



#### Testbed



#### Radio point clouds



#### Signal processing flow

##### 1. OFDM packet echo

Known transmit data  $X(k)$  and received echo  $Y(k)$ .

##### 2. CIR ranging

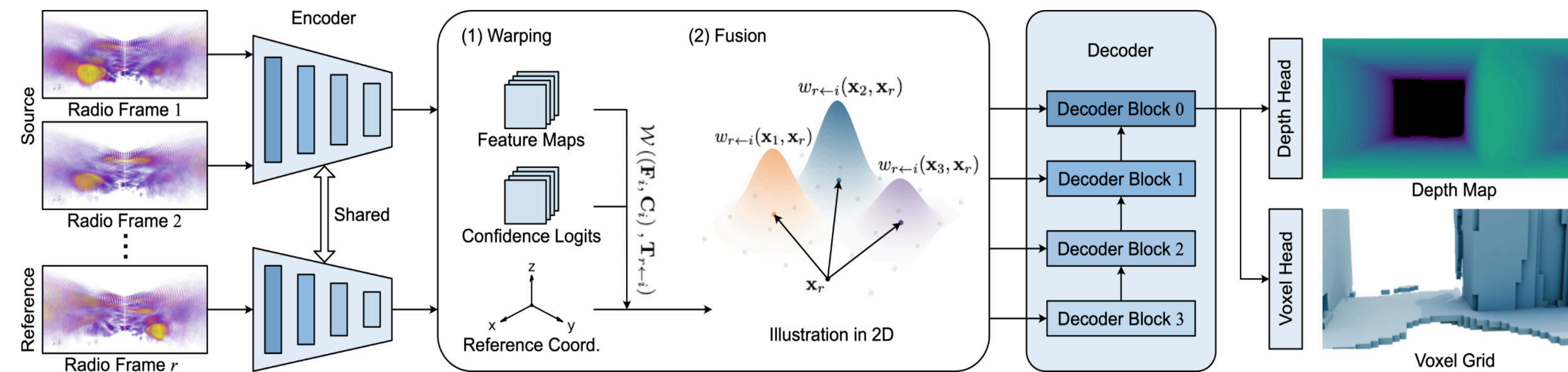
$$\hat{H}(k) = Y(k)/X(k), \hat{h} = \text{IFFT}(\hat{H}), r = nc/(2B).$$

##### 3. Angle scan

Beamform range taps over angles  $(\theta, \phi)$ .

$$\text{Radio frame } S = \{s(n, \theta, \phi)\}$$

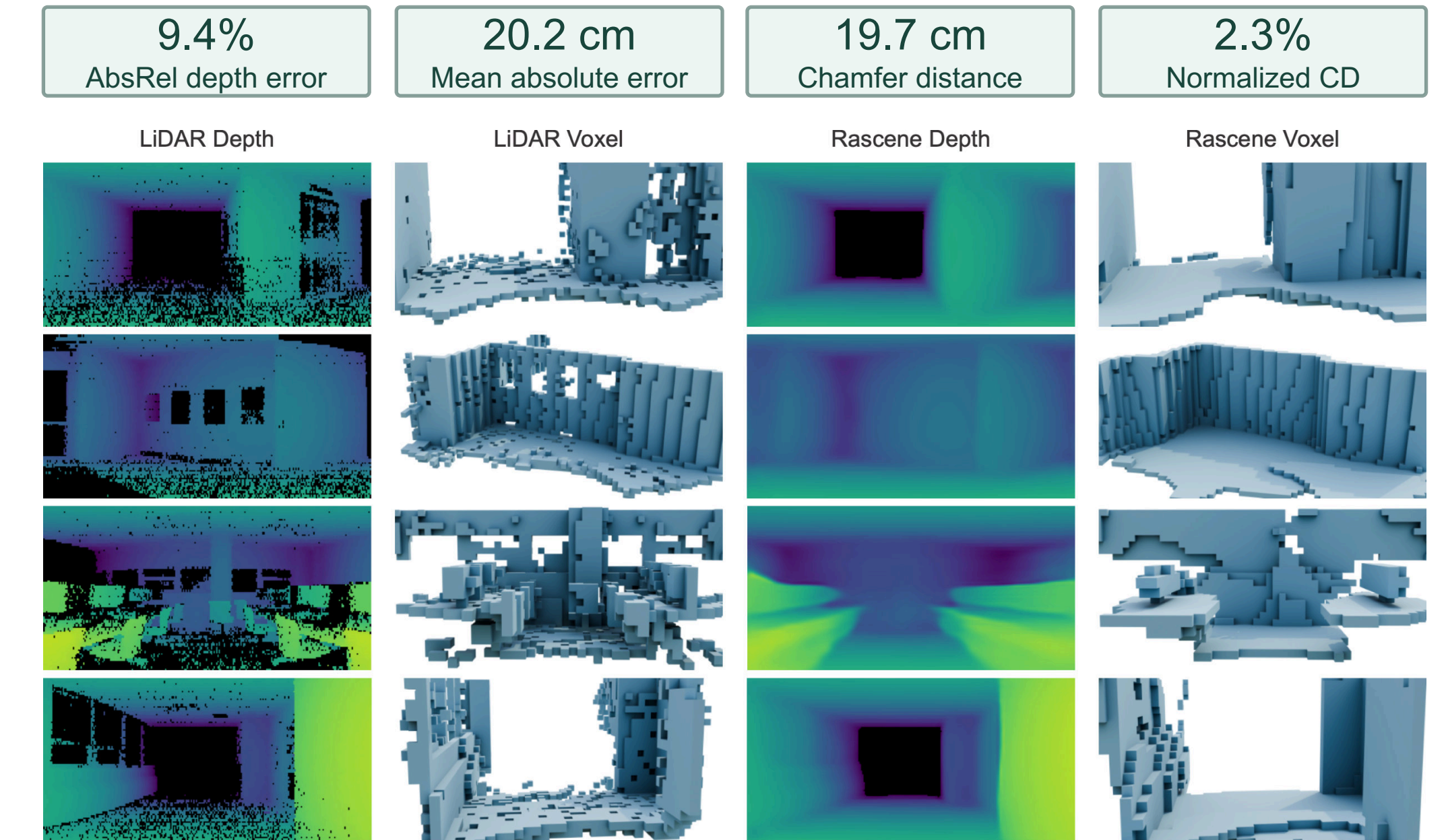
### B. Software: multi-frame 3D RF reconstruction



Given multiple radio frames and poses, a shared encoder predicts per-frame feature maps and confidence logits. We then warp all features to a reference frame and fuse them into a unified representation. A coarse-to-fine 3D decoder with voxel and depth heads outputs the reconstructed voxel grid and depth map.

## 3. Results

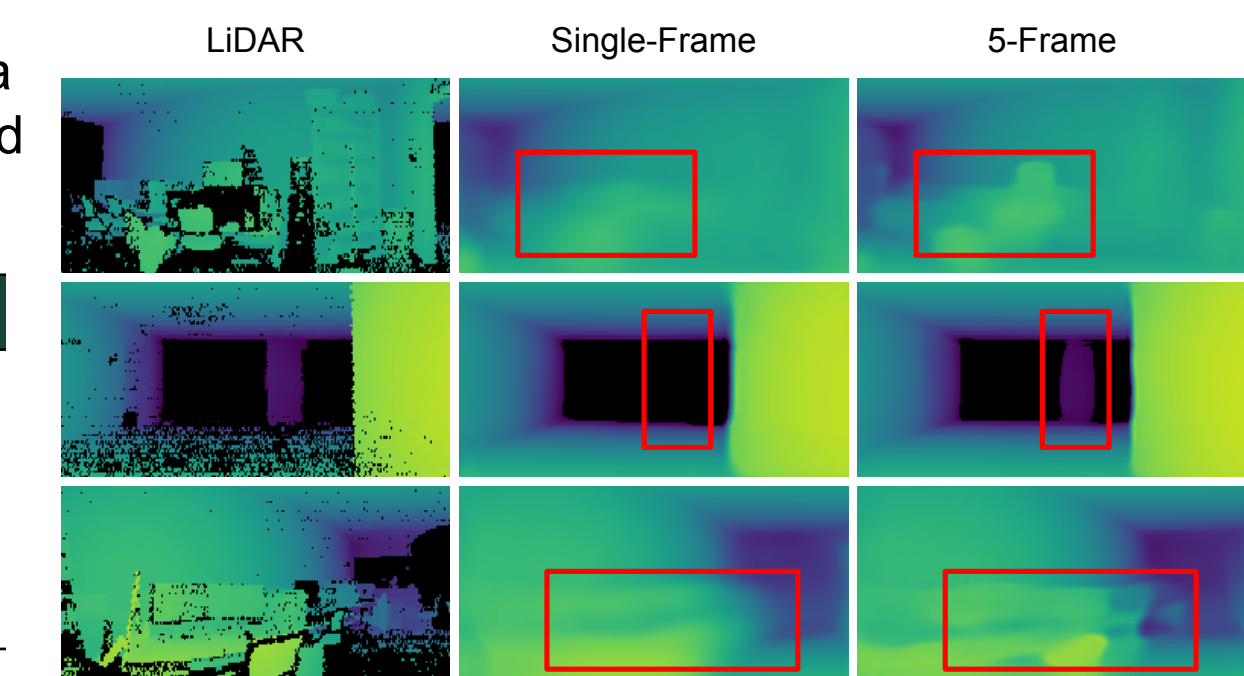
### A. Main results



### B. Multi-frame fusion suppresses ghosts

Each frame observes the scene from a different RF view. Confidence-weighted fusion reinforces consistent geometry.

Frame #	AbsRel	MAE	CD	CD <sub>Diag</sub>
1	14.1%	32.9	31.6	3.6%
2	11.1%	24.6	26.0	3.0%
3	9.8%	21.8	21.9	2.5%
5	9.4%	20.2	19.7	2.3%



### C. Occlusion resilience

Occlusion	AbsRel	MAE	CD	CD <sub>Diag</sub>
No Occ.	5.6%	13.7	18.4	2.0%
Paper Sheet	6.1%	14.4	19.1	2.1%
Styrofoam	6.4%	15.7	21.2	2.3%

