

Institute of Artificial Intelligence

UNIVERSITY OF CENTRAL FLORIDA

FINAL ORAL EXAMINATION

OF

Aakash Kumar

B.E, Sukkur IBA University, 2015
M.S., Shanghai Jiaotong University, 2020

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

(COMPUTER SCIENCE)

26 October 2026, 11:30 A.M.
SPRK 229

DISSERTATION COMMITTEE

Professor Niels Lobo, *Chair*, *Niels.DaVitoriaLobo@ucf.edu*
Professor Chen Chen,, *Co-chair*, *chen.chen@crcv.ucf.edu*
Professor Santu Karmaker, *santu@ucf.edu*
Professor Guifang Li, *li@ucf.edu*

DISSERTATION RESEARCH IMPACT

Three-dimensional perception enables autonomous vehicles, robots, and intelligent environments to understand objects, motion, geometry, and human activity. However, its practical deployment is often constrained by expensive manual annotation, sparse and incomplete geometric measurements, noisy sensor data, and computational limitations. These challenges affect critical applications such as autonomous navigation, human–robot interaction, rehabilitation, assisted living, and privacy-aware monitoring.

This dissertation addresses these challenges through four contributions. It introduces a learned affinity network for LiDAR-based 3D multi-object tracking and a self-supervised framework that reduces the need for manually labeled temporal identities. It also presents a sparse-to-dense reconstruction method that combines images with limited LiDAR measurements to improve 3D object detection. Finally, it develops an efficient mmWave-radar framework for low-latency, privacy-aware 3D human pose estimation. Together, these contributions advance reliable and practical 3D perception under limited supervision, sparse sensing, noisy observations, and real-time deployment constraints.

SELECTED PUBLICATIONS

1. Toward Low-Latency Privacy-Aware 3D Human Pose Estimation with mmWave Radar, A. Kumar, P. Cai, S. H. Musunuri, G.-M. Su, C. Chen, and N. Lobo, submitted to the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2027.

2. Sparse Points to Dense Clouds: Enhancing 3D Detection with Limited LiDAR Data,. A. Kumar, C. Chen, A. Mian, N. Lobo, and M. Shah, in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2024.

3. Self-Supervised Learning for Multiple Object Tracking in 3D Point Clouds, A. Kumar, J. Kini, A. Mian, and M. Shah, in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2022.

4. PC-DAN: Point Cloud Based Deep Affinity Network for 3D Multi-Object Tracking, A. Kumar, J. Kini, A. Mian, and M. Shah in JRDB-ACT Workshop at the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

DISSERTATION

ROBUST 3D PERCEPTION UNDER LIMITED SUPERVISION AND SPARSE SENSING

Progress in three-dimensional perception has been driven by labeled datasets and high-resolution sensors. However, annotating temporal identities for 3D tracking is expensive, while dense geometric sensing requires costly hardware and computation. These limitations motivate methods that learn from limited supervision and sparse or noisy sensor measurements. Privacy is another concern because cameras capture identifiable information. In this dissertation, we present four methods for improving 3D perception under limited supervision, sparse sensing, privacy concerns, and low-latency deployment requirements.

Our first contribution is Point Cloud Based Deep Affinity Network (PC-DAN) for 3D multi-object tracking. Unlike hand-crafted association methods, PC-DAN learns object representations from LiDAR point clouds. A shared PointNet backbone extracts features from detections in consecutive frames, while an affinity network constructs pairwise features and predicts temporal correspondences. This learned formulation establishes the supervised tracking architecture used in our subsequent work.

While PC-DAN provides effective 3D object association, it requires manually annotated temporal identity labels. Our second contribution, SS3D-MOT, introduces self-supervised learning for multiple-object tracking in 3D point clouds. The method applies object-level translation, rotation, and cutout transformations to a single LiDAR frame to generate paired frames with known correspondences. It also generates synthetic scenes using CAD objects and primitive geometric shapes, enabling affinity learning without manually labeled temporal associations.
The preceding contributions address association and supervision but assume that geometric measurements are available. Our third contribution, Sparse Points to Dense Clouds, addresses limited sensing for 3D object detection. A transformer-based framework combines an image with sparse 3D points to reconstruct a dense point cloud. Using only 512 points, approximately one percent of a full KITTI LiDAR frame, the reconstructed geometry consistently improves multiple multimodal 3D detectors.

Our fourth contribution, mmWave-STSFormer, addresses privacy-aware 3D human pose estimation from sparse and noisy mmWave radar point clouds. Unlike cameras, mmWave radar does not capture identifiable visual appearance and remains effective under darkness, smoke, and occlusion. The proposed causal, single-pass framework combines local spatial refinement, point-level temporal fusion, hierarchical token reduction, and global temporal modeling. It achieves robust performance across standard, cross-subject, cross-environment, and adverse-scene settings while reducing computation and latency compared with diffusion-based alternatives. Together, these contributions advance reliable and practical 3D perception under real-world supervision, sensing, privacy, and deployment constraints.



Aakash Kumar

1993	Born in Shikarpur, Sindh, Pakistan
2011-2015	B.E, Sukkur IBA University, Sukkur, Pakistan
2015-2016	Research Assistant, Sukkur IBA University, Pakistan
2016-2018	Visiting Faculty, Sukkur IBA University, Pakistan
2018-2020	M.S., Shanghai Jiaotong University, China
2020-2026	Ph.D., University of Central Florida
2025	Ph.D. Research Intern, Dolby Laboratories, Sunnyvale CA
2026-	Senior Machine Learning Engineer, Intuitive Surgical