Mosam Dabhi

PH.D. STUDENT

Carnegie Mellon University, Pittsburgh, PA, USA

✉ mosam@cmu.edu | ⌂ mosamdabhi.github.io

Research Interests

My PhD research is dedicated to advancing 3D vision and multimodal applications by developing geometric foundation models. These efforts aim to enhance scene understanding and human-computer interaction, pushing the boundaries of general AI, out-of-distribution (OOD) reasoning, and 3D vision graphics. This research lays the foundation for novel approaches in generative AI, robotics – enabling robust applications across a wide range of disciplines.

Computer vision	3D Vision, Multi-view geometry [1–3]
Machine Learning	Optimization, SSL, Generative modeling, Graph [2–4]
Graphics	Implicit Functions, Differentiable Rendering [4]
Robotics	SLAM, Control-theory [5–9]

Education

Carnegie Mellon University *Pittsburgh, USA*

PH.D. IN ROBOTICS *Aug. '21 - Present*

- **Advisor**: Simon Lucey and Laszlo Attila Jeni

Carnegie Mellon University *Pittsburgh, USA*

M.S. IN ROBOTICS *Aug. '19 - May '21*

- **Advisor**: Simon Lucey
- Thesis: Multi-view NRSfM: Affordable Setup for High-Fidelity 3D Reconstruction

National Institute of Technology *Surat, India*

B.TECH. IN ELECTRONICS AND COMMUNICATION ENGINEERING *Aug. '13 - May '17*

Research Experiences

Carnegie Mellon University *Pittsburgh, USA*

PH.D. STUDENT *Aug. '19 - Present*

- **Advisors**: Simon Lucey, Laszlo Attila Jeni
- Set forth the concept of Geometric Foundation Models to enhance 3D vision for AI, reasoning, and action capabilities, focusing on integrating geometry with foundational models.
- Developed and introduced models including MV-NRSfM [1], MBW [2], 3D-LFM [3], VID-3D [4] advancing scalable 3D label generation and ground truth in natural settings.
- MBW enables auto-labeling at scale with minimal supervision, using spatiotemporal neural priors for landmark labeling.
- 3D-LFM presents a universal lifting model from 2D to 3D landmarks, employing graph-based transformers to ensure consistency across diverse object categories without requiring specific object knowledge.
- VID-3D takes RGB videos and generates dense animatable 3D models of deforming objects without pre-defined 3D surface templates, that also enables novel applications like gen-AI video correction.

Apple Inc. *Cupertino, US*

RESEARCH SCIENTIST INTERN (AI) *May '24 - Present*

- LLMs, LVMs, and 3D (*Vision Pro team*).

Apple Inc. *Cupertino, US*

RESEARCH SCIENTIST INTERN (AI) *June '23 - Aug. '23*

- Multimodal time-series modeling, shaping the foundation for lightweight AI foundation models (*Vision Pro team*).

Apple Inc. *Cupertino, US*

RESEARCH SCIENTIST INTERN *May '22 - Aug. '22*

- Few-shot learning and Out-Of-Distribution adaptation algorithms (Vision Pro team).
- Auto-labeling in the wild, driving substantial financial savings by auto-generating 3D labels for computer vision applications .

Apple Inc. *Cupertino, US*

Research Intern *May '21 - Aug. '21*

- Active learning and self-supervised learning strategies (*Vision Pro team*).

Apple Inc. *Cupertino, US*

Research Intern *May '20 - Aug. '20*

- Foundational work on ML based multi-view 3D geometry, paving the way for economical setups to generate 3D groundtruth labels for 3D vision applications (*Vision Pro team*).

Carnegie Mellon University *Pittsburgh, US*

Research Assistant *May '17 - May. '19*

- **Advisor**: Nathan Michael
- Robotic exploration and mapping in real-time for search and rescue operations, enabling superior robot-to-robot communication on extra-terrestrial and sub-terrestrial surfaces [9].
- Achieved flights in challenging, GPS-denied terrains, hitting accelerations over $12 \, \text{m/s}^2$ [8]
- Experience-driven Model Predictive Control (EPC) tailored for platforms with computational constraints [5, 6].
- Motion and path planning in cluttered environments through mixed-integer programming [7].

Indian Institute of Science *Bangalore, India*

Research Intern *May '15 - Jul. '15*

- **Advisor**: Prasanta Kumar Ghosh
- Home automation via speech, harnessing MFCC features and speaking rate analysis.

Publications

[1] **Mosam Dabhi**, Chaoyang Wang, Kunal Saluja, Laszlo Jeni, Ian Fasel, and Simon Lucey. High Fidelity 3D Reconstructions with Limited Physical Views. In *2021 International Conference on 3D Vision (3DV)*. IEEE, 2021.

[2] **Mosam Dabhi**, Chaoyang Wang, Tim Clifford, Laszlo Jeni, Ian Fasel, and Simon Lucey. Multi-view Bootstrapping in the Wild. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*. NeurIPS, 2022.

[3] **Mosam Dabhi**, Laszlo A Jeni, and Simon Lucey. 3d-lfm: Lifting foundation model. *2024 Computer Vision and Pattern Recognition (CVPR)*, 2024.

[4] **Mosam Dabhi**, Simon Lucey, and Laszlo A Jeni. Template-free deformable 3d models from videos, 2024. Manuscript in submission.

[5] **Mosam Dabhi**, Vishnu R Desaraju, and Nathan Michael. Evaluation of Explicit Experience-driven Predictive Control on a Computationally Constrained Platform. Technical report, Carnegie Mellon University, Pittsburgh, PA, June 2017.

[6] **Mosam Dabhi**, Alexander Spitzer, and Nathan Michael. Aggressive Flight Performance using Robust Experience-driven Predictive Control Strategies: Experimentation and Analysis. Technical Report CMU-RI-TR-19-08, Carnegie Mellon University, Pittsburgh, PA, June 2019.

[7] **Mosam Dabhi**, Vishnu Desaraju, and Nathan Michael. Planning Aggressive, Dynamically Feasible and Optimal Trajectories for Autonomous Vehicles in Cluttered Environments using Mixed Integer Programming. Technical report, Carnegie Mellon University, Pittsburgh, PA, 2016.

[8] Alex Spitzer, Xuning Yang, John Yao, Aditya Dhawale, Kshitij Goel, **Mosam Dabhi**, Matt Collins, Curtis Boirum, and Nathan Michael. Fast and agile vision-based flight with teleoperation and collision avoidance on a multirotor. In *International Symposium on Experimental Robotics*, pages 524–535. Springer, 2018.

[9] Wennie Tabib, Kshitij Goel, John Yao, **Mosam Dabhi**, Curtis Boirum, and Nathan Michael. Real-Time Information-Theoretic Exploration with Gaussian Mixture Model Maps. In *Robotics: Science and Systems*, 2019.

Honors & Awards

'19 - '23	**Apple Research Grant**, Apple Inc.
'17	**Research Scholarship**, FICCI, India
'16 - '17	**Summer Scholar**, Robotics Institute Summer Scholar
'16	**Undergraduate thesis funding**, TEQIP Award, MHRD, Government of India

Academic Services

'22 - '24	**Conference Paper Reviewer**, NeurIPS; CVPR; ICCV; ECCV; WACV
'21 - '22	**Conference Paper Reviewer**, ICRA, IROS
'20	**Conference Paper Reviewer**, International Conference on Humanoid Robots
'21 - '24	**Admissions Committee**, Master of Science, Robotics, CMU (MSR)
'23 - '24	**Admissions Committee**, Master of Science, Computer Vision, CMU (MSCV)
'22 - '23	**M.S. in Robotics Thesis Committee** , Examinee: Heng Yu, Aarush Gupta
'17 - '19	**Admissions & Admin. Committee**, Robotics Institute Summer Scholars, CMU (RISS)

Teaching Experiences

Carnegie Mellon University

TEACHING ASSISTANT

- **Spring 2022** : Robot Localization and Mapping with Prof. Michael Kaess
- **Fall 2022**   : Geometry-Based Methods in Vision with Prof. Shubham Tulsiani

Relevant coursework

Carnegie Mellon University

LEARNING FOR 3D VISION (**A+**), ADVANCED COMPUTER VISION (**A**), GEOMETRY METHODS IN VISION (**A+**), ADVANCED MACHINE LEARNING (**A**), CONVEX OPTIMIZATION (**A**), ROBOT LOCALIZATION AND MAPPING / SLAM (**A+**), MATH. FUNDAMENTALS FOR ROBOTICS (**A**), KINEMATICS, DYNAMICS, AND CONTROLS (**A+**)

Proficient Skills

Programming languages

PRIMARY: PYTHON, C/C++, LaTeX, MATLAB

SECONDARY: CUDA, LUA, HTML, JAVASCRIPT

Software libraries

PRIMARY: PYTORCH, TENSORFLOW, BLENDER, COLMAP

SECONDARY: OPENAIGYM, TORCH, CAFFE, OPENCV, VLFEAT, PTHREAD