# Siddharth Choudhary

itzsid.github.io

Google Scholar: 1,223 citations, h-index: 14

Email : sidchoud@amazon.com

Mobile : +1-404-519-2452

## EDUCATION

**Georgia Institute of Technology** — Atlanta, GA
*Ph.D. in Computer Science; GPA: 3.84* — *Aug. 2012 – Aug 2017*

**IIIT Hyderabad** — Hyderabad, India
*Master of Science in Computer Science; GPA: 4.0 (10.0/10.0)* — *Aug. 2010 – July. 2012*

**IIIT Hyderabad** — Hyderabad, India
*Bachelor of Technology (Honors) in Computer Science; GPA: 3.31 (8.28/10.0)* — *Aug. 2006 – July. 2010*

## EXPERIENCE

**Amazon AGI** — San Francisco Bay Area, CA
*Principal Applied Scientist* — *April 2025-Present*

- **Amazon Nova 2.0 Lite**: Tech-lead responsible for designing the architecture and finalizing the pretraining data mix and recipe for Nova 2.0 Lite model integrating vision, speech and language understanding capabilities at scale. Launched at AWS re:Invent 2025. Achieves state-of-the-art performance on 13 out of 15 benchmarks as compared to Claude Haiku 4.5.
- **Amazon Nova 2.0 Omni**: Led development of unified multimodal architecture integrating native image generation with text, vision, and speech understanding. Designed pre-training recipe that enable generative capabilities while preserving performance across all input modalities. Launched at AWS re:Invent 2025.

*Senior Applied Scientist* — *January 2024-April 2025*

- **Amazon Nova 1.0 Foundation Models**: Tech-lead on multimodal pre-training team responsible for developing image/video pretraining recipes for Amazon Nova foundation models launched at AWS re:Invent 2024. Achieved state-of-the-art performance on benchmarks against industry leaders Claude 3 Haiku and Gemini 1.5 Pro on multimodal understanding benchmarks.

**Amazon Web Services (AWS)** — San Francisco Bay Area, CA
*Senior Applied Scientist* — *May 2023-January 2024*

- **Large vision language model**: Developed state-of-the-art multimodal model combining 1.8B ViT with a 100B parameter language model, outperforming SOTA open source models across 12 benchmarks. Designed 3-stage pretraining recipe within 1.5-month timeline.
- **Multimodal Hallucination Control**: Developed M3ID sampling method reducing VLM hallucinations by up to 28% without additional training. Published at CVPR 2024 and featured in AWS keynote at CVPR. Mentored intern on visual information grounding techniques.

**Amazon Lab126** — Sunnyvale, CA
*Senior Applied Scientist* — *August 2020-May 2023*

- **Halo Body**: Developed MeasureNet CNN model for body measurements and waist-hip ratio prediction from smartphone images, achieving twice the accuracy of state-of-the-art models. Uniquely trained using only synthetic data, eliminating privacy concerns. Published in Nature Digital Medicine (2023) and filed multiple patents.
- **Halo Trainer**: Built transformer-based models for fitness activity understanding, including on-device real-time repetition counting and form error detection. Won Best Paper Award at Amazon Computer Vision Conference 2023.

**Magic Leap** — Sunnyvale, CA
*Principal Computer Vision Researcher/Engineer* — *October 2017-September 2020*

- **3D Object Recognition in AR Cloud**: Lead architect and developer for multiuser, scalable 3D object recognition system deployed in Magic Leap One (ML-19). Built complete pipeline from scratch with low memory footprint scaling to large numbers of concurrent users. Published at CVPR 2020 AR/VR Workshop and filed patents.
- **Learned Keyframe Selection for SLAM**: Designed PointNet-based neural network for frame embedding to improve localization recall by 10-30% while reducing memory requirements by 75% compared to Bag of Words algorithm.

- **Scalable Infrastructure for SLAM Research**: Built scalable infrastructure to extract data from SLAM pipeline stages with ground-truth, enabling training and evaluation of machine learning algorithms for SLAM research.

- **Institute for Robotics and Intelligent Machines, Georgia Tech** — Atlanta, GA
  *Graduate Research Assistant* — *August 2012-August 2017*
  - **Distributed Object-based SLAM**: Developed distributed algorithms using Distributed Gauss-Seidel methods for multi-robot trajectory estimation with minimal information exchange. Extended framework to include object-level semantics for distributed object-based SLAM. Published extensively in IJRR, ICRA, IROS with more than 500 citations.
  - **Memory-efficient SLAM**: Proposed exactly sparse SLAM approach using multi-block Alternating Direction Method of Multipliers (ADMM) to enforce consistency among subgraphs.

- **Center for Visual Information Technology, IIIT Hyderabad** — Hyderabad, India
  *Research Assistant* — *August 2010-August 2012*
  - **Bundle Adjustment on GPU**: Developed hybrid CPU-GPU implementations of sparse bundle adjustment achieving 30-40x speedup over standard CPU implementations on datasets with up to 500 images using NVIDIA Tesla C2050 GPU. Published in ECCV 2010 workshop on computer vision on GPUs.
  - **Visibility Probability Structure from SfM Datasets**: Developed visibility probability structures encoding visibility information between points and cameras as conditional probabilities for improved image localization. Published at ECCV 2012 with 86+ citations.

## PUBLICATIONS

**Multimodal LLMs**
- The Amazon Nova Family of Models: Technical Report and Model Card (**AWS re:Invent 2024 (38 citations)**)
- Multi-modal hallucination control by visual information grounding (**CVPR 2024 (157 citations)**)
- RAVEN: Multitask Retrieval Augmented Vision-Language Learning (**arXiv 2024 (7 citations)**)

**Computer Vision & AR**
- SplatArmor: Articulated Gaussian splatting for animatable humans from monocular RGB videos (**arXiv 2023 (45 citations)**)
- Development and Validation of an Accurate Smartphone Application for Measuring Waist-to-Hip Circumference Ratio (**Nature Digital Medicine 2023 (6 citations)**)
- Mesh Strikes Back: Fast and Efficient Human Reconstruction from RGB videos (**arXiv 2023**)
- Multi-task Transformer for Real-Time Fitness Activity Understanding from Videos (**Amazon Computer Vision Conference 2023 (Best Paper Award)**)
- Multiuser, Scalable 3D Object Detection in the AR Cloud (**CVPR 2020 workshop on AR/VR (4 citations)**)
- Practical time bundle adjustment for 3d reconstruction on the gpu (**ECCV 2010 Workshop (51 citations)**)
- Visibility Probability Structure from SfM Datasets and Applications (**ECCV 2012 (86 citations)**)
- CPU and/or GPU: Revisiting the GPU vs. CPU myth (**arXiv 2013 (16 citations)**)

**Robotics & SLAM**
- Data-efficient decentralized visual SLAM (**ICRA 2018 (279 citations)**)
- Distributed Mapping with Privacy and Communication Constraints: Lightweight Algorithms and Object-based Models (**IJRR 2017 (160 citations)**)
- SLAM with Object Discovery, Modeling and Mapping (**IROS 2014 (66 citations)**)
- Information-based Reduced Landmark SLAM (**ICRA 2015 (64 citations)**)
- Distributed Trajectory Estimation with Privacy and Communication Constraints: a Two-Stage Distributed Gauss-Seidel Approach (**ICRA 2016 (57 citations)**)
- Multi Robot Object-based SLAM (**ISER 2016 (47 citations)**)
- Exactly Sparse Memory Efficient SLAM using the Multi-Block Alternating Direction Method of Multipliers (**IROS 2015 (32 citations)**)

## PATENTS

- Body dimensions from two-dimensional body images (**US Patent 11,861,860 (2024) (6 citations)**)
- Cross Reality System (**US Patent 11,227,435 (2022) (83 citations)**)
- Scalable three-dimensional object recognition in a cross reality system (**US Patent 11,257,300 (2022) (4 citations)**)
- Scene understanding using occupancy grids (**US Patent 12,444,136 (2022)**)

## Academic Duties

- Conference reviewer for CVPR, ICCV, ECCV, 3DV since 2020.

- Conference reviewer for ICRA, IROS since 2014

- Journal reviewer for IEEE Transactions of Robotics.

- Journal reviewer for IEEE Robotics and Automations letters.

- Journal reviewer for International Journal for Robotics Research.

- Program Committee member of Robotics Science and Systems conference in 2016.

- Member of the Organizing committee for the tutorial on Perception at Magic Leap at the 2019 Conference on Computer Vision and Pattern Recognition.

## Programming Skills

- **Languages**: Python, CUDA, Shell, JavaScript, C++

- **ML/AI Frameworks**: PyTorch, Megatron-LM, Transformers, DeepSpeed, Numpy, SciPy

- **Cloud/Infrastructure**: AWS SDK, Docker, Kubernetes