

Venkata Harshavardhan Bontalakoti

✉ vbontal1@jh.edu | 📞 (+1) 410-908-8940 | [GitHub](#) | [LinkedIn-Harshavardhan](#) | [Portfolio](#)

EDUCATION

Johns Hopkins University , Baltimore, MD, U.S.A M.S.E. in Robotics Engineering GPA: 3.75/4.0	08/2025 – 05/2027
Indian Institute of Technology (IIT) Hyderabad & NIT Silchar , India B.Tech (Robotics & Autonomous Systems Focus) GPA: 9.20/10.0 (IIT-H Exchange, Top 10%)	08/2021 – 05/2025

WORK EXPERIENCE

Capstan Medical , California, U.S.A (Robotics Software Intern)	06/2026 – 08/2026
- Built a custom catheter physics simulator modeling bending, roll, and insertion kinematics coupled to the production robot controller, enabling closed-loop controller testing and failure-case exploration without physical hardware time. - Implemented power-on self-test (POST) routines in C++ for range-of-motion and stiction checks within a regulated surgical robotics codebase with comprehensive unit tests. Characterized actuator and robot motion by writing automated analysis pipelines over motion logs to quantify friction and range limits, defining pass-fail thresholds for the production POST suite.	
Johns Hopkins University , Maryland, U.S.A (Graduate Researcher & Spark Fellow)	09/2025 – 06/2026
- Engineered GenFaD, a generalized visual defect and crack-detection framework for autonomous railway inspection combining classical morphological filtering with deep semantic segmentation models reaching 99.73% accuracy on in-domain test data. - Developed computer vision and AI perception models for public safety analytics and real-time environmental monitoring.	
Defense Research and Development Organization (DRDO) , India (Robotics Intern)	05/2023 – 07/2023
Developed a multi-sensor hardware-software integration pipeline for EO/IR cameras and 3D LiDAR payloads on mobile defense platforms, and prototyped a distributed buoy coastal surveillance network reducing projected operational costs by 15%.	

PROJECTS

Multi-Task Manipulation with Interpretable Vision-Language-Action (VLA) Head <i>PyTorch, Vision-Language-Action Models (VLA), Multi-GPU Distributed Training</i>	<i>Ongoing Research</i>
Designing and benchmarking a novel lightweight spatial-dynamics action head for Vision-Language-Action (VLA) policies to improve parameter efficiency and interpretability in multi-task robotic manipulation. Achieved a mean success rate of 90.2% on LIBERO benchmark on a frozen vision-language backbone.	
Dreaming of Birds: Camera-to-BEV Map Generation with Diffusion Transformers <i>PyTorch, Diffusion Transformers (DiT), DINOv2, 3D Ray Projection, nuScenes, Multi-View Computer Vision</i>	[GitHub]
Built a Diffusion Transformer (DiT) in PyTorch that predicts 6-layer Bird's-Eye-View (BEV) semantic maps (roads, crossings, lane dividers, vehicles, pedestrians) from 6 surround cameras on nuScenes; cross-attended to frozen DINOv2 features tagged with 3D camera viewing rays, achieving 0.319 mIoU vs. 0.260 for a U-Net baseline.	
ClipSort: Zero-Shot Multi-Object Tracking with Foundation Model Embeddings <i>Python, PyTorch, YOLOv11, CLIP, DINOv2, OpenCV, Kalman Filtering, Hungarian Algorithm</i>	[GitHub]
Architected a modular Multi-Object Tracking (MOT) framework combining YOLOv11, Kalman Filter state estimation, and Hungarian data association; replaced task-specific Re-ID CNNs with zero-shot CLIP, DINOv2, and ResNet visual embeddings to benchmark identity-preservation vs. inference latency trade-offs under heavy occlusion.	
CERBRUS: Vision-Based 12-DOF Quadruped Robot with Sensor Fusion <i>ROS 2, C++, Python, MuJoCo, Extended Kalman Filter (EKF), Visual-Inertial Odometry, Pose Estimation</i>	[GitHub]
Built a 12-DOF open-source quadruped with a 2-DOF gripper, implementing real-time pose estimation (MoveNet) and IMU-camera sensor fusion (EKF) in ROS 2 and MuJoCo for collaborative aerial-ground autonomous navigation.	

ACHIEVEMENTS

- YCP Hacks 2025 (2nd Place Overall & Hardware Winner):** Built *PerceptEye*, an edge perception system on Raspberry Pi combining real-time face recognition, sign-language detection, and object detection for visually/hearing-impaired navigation.
- NexHacks CMU 2026 (Sponsor Track Winner):** Engineered an AI-powered Smart Medical Kit using a multi-agent LLM consensus architecture for automated first-aid triage and diagnostic reasoning.
- Smart India Hackathon 2024 Winner (Hardware Edition):** Won 1st place among 100+ teams by engineering an autonomous drone-based aerial monitoring and smart automation system.

PUBLICATIONS

- Autonomous aerial deployment of a hybrid drone-quadruped system for environmental tasks**, *Proc. IMechE Part C: Journal of Mechanical Engineering Science* (2026). [\[Paper\]](#)
- Improvement in Multi-resident Activity Recognition System in a Smart Home Using Activity Clustering**, *IFIP IoT* (2023). [\[Paper\]](#)

TECHNICAL SKILLS

Computer Vision & Embodied AI: PyTorch, OpenCV, Vision-Language-Action Models (VLA), Foundation Models (CLIP, DINO), Multi-Object Tracking (DeepSORT, YOLOv11), 3D Vision & Depth Estimation, Sensor Fusion (EKF), TensorFlow

Robotics & Simulation: ROS 2, MuJoCo, Unity, Gazebo, SLAM, Nav2, MoveIt 2, MPC, Trajectory Planning

Programming & Tools: C++ (C++17, STL, Unit Testing), Python, C, Linux, Git, Docker, MATLAB, C#