Trung M. Bui, PhD

Robotics & Computer Vision Engineer — Robot Learning, Manipulation & Perception

Morgan Hill, CA • +1 669-326-2460 • bmtrungvp@gmail.com
linkedin.com/in/trung-m-bui • github.com/mtbui2010 • aistations.org • **U.S. Green Card**

Summary

Robotics and Computer Vision Engineer (PhD) with 7+ years shipping production perception and 11+ years total CV experience, focused on **foundation models for dexterous manipulation**. Builds and evaluates **vision-language-action (VLA) policies** in large-scale simulation (**Isaac Lab, MuJoCo**) — reaching **98.75% on the LIBERO benchmark** with OpenVLA-OFT — and uses **RL fine-tuning and diffusion policies** to cross the embodiment gap from sim to real. Extracts **3D structure** (6D pose, RGB-D, point clouds) and grounds it with open-vocabulary vision models (GroundingDINO, CLIP, FastSAM) for **multimodal sensor fusion**. Took an industrial manipulation system from **70% to 92% success in production**, **30+ FPS on Jetson** via TensorRT/INT8, deployed 3+ years at scale. Published in *IEEE TIP* (Q1) and *IEICE Transactions*.

Experience

Senior Computer Vision, Robotics & AI Engineer *March 2019 – Present*

Korea Electronics Technology Institute (KETI), Seongnam, South Korea

- Led perception ML for a **production manipulation / bin-picking system**: improved grasp success from 70% to **92% in cluttered real-world scenes**, processing hundreds of items per hour, deployed 3+ years across multiple sites at scale.
- Trained and shipped models for **object detection, segmentation, and 6D pose estimation** on RGB-D data — extracting 3D structure for grasping; handled the long tail via active learning and open-vocabulary detection.
- Designed a **vision transformer with fine-grained attention** for precision grasp-pose detection — **20% precision improvement over CNN baselines**. Published in *IEICE Transactions*, Nov 2025.
- **Architected MARS (Mobile Assistive Robotic System)** — a 7-DOF mobile manipulator with hybrid gripper and head/wrist RGB-D on a **4-layer ROS2 architecture**: Task Manager (LLM/VLM planning), hierarchical Skill Controller with bounded retries, Hardware Controllers. Published in *Ubiquitous Robots 2025*.
- Integrated **VLMs (CLIP, BLIP, LLaVA, GroundingDINO, Gemini Vision)** for zero-shot / open-vocabulary recognition and semantic grounding fused with depth for manipulation targets.
- Built end-to-end **training pipelines**: dataset curation (synthetic + real + production failures), augmentation, evaluation harnesses, and continuous improvement loops — closing the loop from real-world data back into the models.
- Optimized inference for edge with **TensorRT, ONNX, INT8 quantization** on NVIDIA Jetson — **30+ FPS**, 50% latency reduction, sub-200ms perception-to-action latency.
- Led cross-functional collaboration across perception, robotics, hardware, and deployment teams; mentored 3 junior engineers. Contributed core components to national R&D projects with multi-million USD funding.

Selected Projects

VLA Manipulation Policies on Franka Panda (Isaac Lab + MuJoCo) *2025*

github.com/mtbui2010/robot_sim_vla

- Built a **vision-language-action (VLA)** evaluation platform for a Franka Panda arm across **Isaac Lab** (Isaac Sim 4.5) and **MuJoCo** with standardized observation/action spaces — large-scale simulation infrastructure for manipulation foundation models.
- Evaluated pretrained **OpenVLA-OFT** on **LIBERO**: **98.75% overall success** (80 episodes; 100% Spatial/Object/Goal, 95% Long), edging the 97.1% paper baseline with only 2 trials/task.
- Reproducible fine-tuning infra (three isolated Conda envs, 24–30h on 4×A6000 / 150K steps) and a **LeRobot** diffusion-policy baseline conversion pipeline for sim-to-real transfer.

MARS — Mobile Assistive Robotic System *Ubiquitous Robots 2025*

keti-ai.github.io/carerobotdocs | robotapp | kcare_robot

- 7-DOF arm on a 2-wheeled mobile base, hybrid gripper (parallel + suction), head + wrist RGB-D, on a 4-layer ROS2 framework.
- LLM/VLM task planning grounded in skill affordances with ambiguity resolution; Skill Controller with primitive skills (Move, Find, Pick, Place), precondition checks, recursive fallback, n=2 retry bound.
- Modular packages: `pyconnect`, `pyrecognition` (grasp / detection / segmentation / VLM), `pyinterfaces`, `rosinterfaces`.

Tracker Lab — Multi-Object Tracking from Scratch *2025*

github.com/mtbui2010/vision_tracking

- Implemented **SORT, DeepSORT, ByteTrack, and a custom tracker from scratch** — Kalman filter, Hungarian

assignment, motion/appearance/IoU cost functions — with unit tests and derivations, no third-party tracking libraries.
- Detection via fine-tuned **YOLOv11** (MOT17/DanceTrack); ReID via **OSNet**; custom **MOTA** / **IDF1** / **HOTA** metrics benchmarked against MOT17-val.

VisionServe — Local-First CV Inference Server ("Ollama for Computer Vision") *2025*

github.com/mtbui2010/vision_serve

- **Single-binary Go inference server** serving **16 vision models** (RF-DETR, MobileSAM / SAM2, GroundingDINO, Depth Anything V2, CLIP, OCR) behind one unified REST API — a real-time perception backend for robots.
- **ONNX Runtime with multi-backend execution providers** and automatic fallback (**TensorRT** → **CUDA** → **CPU**); edge-first, optimized for **NVIDIA Jetson**, Docker variants for GPU / CPU / ARM64.
- Production serving patterns: **lazy load with idle auto-unload**, thread-safe concurrent sessions, low-latency inference (CLIP 33 ms, RF-DETR-nano 57 ms on RTX A6000), open-vocabulary text prompts.

C++ ML System — ML/CV/Robotics Algorithms from Scratch *2025*

github.com/mtbui2010/cpp-ml-system

- Implemented ML, CV, and robotics primitives in C++17 without frameworks: matrix ops (naive → cache-blocked → tiled → **AVX2+FMA**), im2col+GEMM convolution, tiny CNN (ReLU/maxpool/softmax), NMS, Kalman/EKF, PID with anti-windup, point cloud utilities.
- N-dimensional Tensor with broadcasting; custom memory arena allocator; full **GoogleTest** suite and performance benchmarks isolating the gain at each optimization step.

Single Image Dehazing Using Color Ellipsoid Prior *IEEE TIP 2018*

github.com/mtbui2010/Single-Image-Dehazing

- Novel single-image dehazing algorithm (PhD research); published in *IEEE TIP* (Q1), outperforming existing methods on standard benchmarks.

Selected Publications

- **T. M. Bui**, J. Hwang, S. Jun, W. Kim, D. Shin. "A Fine-Aware Vision Transformer for Precision Grasp Pose Detection." *IEICE Transactions on Information and Systems*, Nov. 2025.
- **T. M. Bui**, Y. Kim, S. J. Moon, M. Cho, M. Seo, D. Shin. "Development of a Mobile Assistive Robot for Daily Living Support." *Ubiquitous Robots*, 2025.
- **T. M. Bui**, W. Kim. "Single Image Dehazing Using Color Ellipsoid Prior." *IEEE Transactions on Image Processing*, Feb. 2018. *(Q1 journal)*
- **T. M. Bui**, H. N. Tran, W. Kim, S. Kim. "Segmenting Dark Channel Prior in Single Image Dehazing." *IET Electronics Letters*, March 2014.

Education

Ph.D. in Computer Vision *2014 – 2019*

Kyung Hee University, South Korea

Thesis: *Single Image Dehazing Using Color Ellipsoid Prior* (basis of IEEE TIP 2018 publication)

M.Eng. in Computer Vision *2011 – 2014*

Kyung Hee University, South Korea

B.Eng. in Electrical and Electronics Engineering *2005 – 2010*

Ho Chi Minh City University of Technology, Vietnam — Excellent Engineer Training Program

Technical Skills

Robot Learning	vision-language-action (OpenVLA-OFT), diffusion policies, imitation & RL fine-tuning, sim-to-real / real2sim2real, LIBERO benchmark, LeRobot
Manipulation	dexterous grasping, bin-picking, grasp planning, 6D pose, skill controllers with hierarchical retry, ROS2 / ROS, MoveIt2, hand-eye calibration
3D Perception	detection, semantic / instance segmentation, 6D pose estimation, RGB-D, point clouds, multi-object tracking (SORT / DeepSORT / ByteTrack), scene understanding
Multimodal / Foundation Models	VLMs (CLIP, BLIP, LLaVA, GroundingDINO, FastSAM, Gemini Vision), open-vocabulary detection, sensor fusion, LLM planning, RAG, LangGraph
Simulation	Isaac Lab / Isaac Sim, MuJoCo, AI2-THOR, large-scale evaluation harnesses
Deep Learning	vision transformers, CNNs, attention, transfer / self-supervised learning, hyperparameter tuning
Frameworks	PyTorch, TensorFlow, Hugging Face, ONNX, TensorRT, CUDA, LeRobot
Languages	Python (expert), C++ (C++17), C, Go, TypeScript
Edge Deployment	TensorRT, INT8 quantization, NVIDIA Jetson, real-time inference (30+ FPS), sub-200ms latency
Production ML	training pipelines, dataset curation, augmentation, evaluation harnesses, shadow deployment, canary rollout, drift detection
Infrastructure	Docker, Linux, Git, distributed training, FastAPI, RunPod