Trung M. Bui, PhD

Robotics & Computer Vision Engineer — Embodied AI Systems

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

Summary

Robotics and Computer Vision Engineer with 7+ years building production end-to-end systems — not only perception models, but the full software stack around them. Architected a mobile manipulation system (MARS) with ROS2 framework, LLM-based task planning, and hierarchical skill controller with bounded retries. Led perception for an industrial bin-picking system at 92% success in production, 30+ FPS on Jetson, deployed for 3+ years. Published in *IEEE TIP* (Q1), *IEICE Transactions*, and *Ubiquitous Robots 2025*. Experience integrating VLMs and LLMs with perception pipelines, and evaluating **vision-language-action (VLA)** policies (OpenVLA-OFT) in Isaac Lab and MuJoCo — 98.75% success on the LIBERO manipulation benchmark.

Experience

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

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

- **Architected MARS (Mobile Assistive Robotic System)** — a 7-DOF mobile manipulator with hybrid gripper, RGB-D cameras, ROS2 framework, and LLM-based task planning. Designed **4-layer system architecture**: WebRTC frontend, Task Manager, hierarchical Skill Controller, Hardware Controllers. Published in *Ubiquitous Robots 2025*.
- Built **hierarchical skill execution** (Move / Find / Pick / Place) with bounded retry mechanism (n=2 per level) for predictable execution times and graceful degradation in cluttered environments.
- Led perception for production **bin-picking system**: improved pick success from 70% to **92% in cluttered scenes**, 30+ FPS on NVIDIA Jetson AGX with TensorRT, system in production 3+ years.
- Designed **vision transformer for grasp pose detection** with fine-grained attention — 20% precision improvement over CNN baselines. Published in *IEICE Transactions*, Nov 2025.
- Integrated **VLMs (GroundingDINO, CLIP, BLIP, LLaVA, Gemini Vision)** for open-vocabulary detection and semantic grounding in unstructured environments.
- Built **LLM-based task planning** with structured outputs grounded in robot skill affordances. Implemented multi-agent framework with LangGraph supervisor, episodic + semantic memory (SQLite + ChromaDB).
- Optimized inference with **TensorRT, ONNX, INT8 quantization** on NVIDIA Jetson, achieving sub-200ms perception-to-action latency end-to-end.
- Led cross-functional collaboration across perception, robotics, hardware, and customer deployment teams; mentored 3 junior engineers on production CV code.

Selected Projects

MARS — Mobile Assistive Robotic System *Ubiquitous Robots 2025*

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

- **Hardware:** 7-DOF arm on 2-wheeled mobile base, hybrid gripper (parallel + suction), head + wrist RGB-D.
- **Software architecture:** 4-layer ROS2 framework — Frontend (WebRTC, voice + text), Task Manager (LLM planning), Skill Controller (primitive skills with hierarchical retry), Hardware Controllers.
- **LLM task planning:** structured outputs grounded in robot skill affordances; ambiguity resolution; interactive Confirm / Skip / Terminate loops.
- **Skill controller:** primitive skills (Move, Find, Pick, Place) with precondition checks, recursive fallback, n=2 retry bound for predictable execution.
- **Modular packages:** `pyconnect`, `pyrecognition` (grasp / detection / segmentation / VLM), `pyinterfaces`, `rosinterfaces`.

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 spanning two physics simulators — **Isaac Lab** (Isaac Sim 4.5) and **MuJoCo** — with standardized observation/action spaces and task wrappers.

- Evaluated pretrained **OpenVLA-OFT** policies 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.
- Built reproducible fine-tuning infrastructure (three isolated Conda envs, 24–30h on 4×A6000 for 150K steps) and a **LeRobot** diffusion-policy baseline conversion pipeline.

CareRobotAgent — Multi-Agent Framework with LangGraph *2024 – Present*

github.com/mtbui2010/carerobotagent | pyplanner

- Multi-agent framework: LangGraph supervisor routing + pluggable planner backends (ReAct, CoT, Hierarchical, Self-Refine) via the `pyplanner` library I built.
- Persistent memory: episodic (SQLite) + semantic (ChromaDB) for personalization. Multi-backend LLM (Ollama / OpenAI / Anthropic). Voice via Whisper + gTTS. AI2-THOR simulator integration.

Fine-Aware Vision Transformer for Grasp Detection *Published, IEICE 2025*

- Novel transformer architecture with fine-grained attention module for cluttered RGB-D scenes.
- Achieved 20% precision improvement over CNN baselines on grasp pose detection.
- Deployed on NVIDIA Jetson with TensorRT, sustained 30+ FPS at production accuracy.

GroundingDINO Annotation & Training Tool *2024*

github.com/mtbui2010/groundingdino_tool

- Streamlit-based web tool for semi-automatic image labeling and open-vocabulary detector fine-tuning.
- Pipeline: YOLOv8n for fast initial detection GroundingDINO open-vocabulary refinement FastSAM for fine segmentation human verification fine-tuning loop.
- Designed for the long-tail problem: closes label gaps for novel objects without requiring large pre-labeled datasets.

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, ReID cascade); fine-tuned YOLOv11 detector + OSNet ReID.
- Custom MOTA / IDF1 / HOTA metrics; full-stack platform (FastAPI + Next.js 14) deployed on Docker, Vercel, and RunPod serverless GPU workers.

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

github.com/mtbui2010/Single-Image-Dehazing

- Novel algorithm for single-image dehazing; PhD research basis.
- Published in *IEEE Transactions on Image Processing* (Q1 journal).
- Outperformed existing methods on standard benchmarks across diverse atmospheric conditions.

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

Robotics: mobile manipulation, bin-picking, grasp planning, skill controllers with hierarchical retry, ROS2 / ROS, MoveIt2, hand-eye calibration
Computer Vision: object detection, segmentation, 6D pose estimation, RGB-D, point clouds, scene understanding
Embodied AI / Multimodal: vision-language-action (VLA) policies (OpenVLA-OFT), VLMs (GroundingDINO, CLIP, BLIP, LLaVA, Gemini Vision), LLM-based planning, RAG, prompt engineering, agent frameworks (LangGraph)
Simulation: Isaac Lab / Isaac Sim, MuJoCo, AI2-THOR, LeRobot, LIBERO benchmark
Deep Learning: vision transformers, CNN architectures, attention, transfer learning, model adaptation
Frameworks: PyTorch, TensorFlow, Hugging Face, ONNX, TensorRT, CUDA
Languages: Python (expert), C++, C
Edge Deployment: TensorRT, INT8 quantization, NVIDIA Jetson, real-time inference (30+ FPS), sub-200ms latency
Production: training pipelines, evaluation harnesses, shadow deployment, canary rollout, telemetry design
Infrastructure: Docker, Linux, Git, distributed training, Streamlit, FastAPI
Certifications: Visual Perception for Self-Driving Cars (Coursera, 2023); Improving Deep Neural Networks (DeepLearning.AI, 2021)