Trung M. Bui, PhD

Robotics, Computer Vision & Multimodal AI Engineer — Production & Embodied AI Systems

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

Computer Vision, Robotics, and Multimodal AI Engineer with 7+ years shipping production systems and 11+ years total CV experience including PhD research. Builds the full stack around perception models — detection, segmentation, 6D pose estimation, multi-object tracking — and integrates **VLMs, LLMs, and vision-language-action (VLA) policies** into grounded, multi-step robot execution. Architected a mobile manipulation system (MARS, *Ubiquitous Robots 2025*) and led an industrial bin-picking system to **92% success in production, 30+ FPS on Jetson, 3+ years deployed**. Recent work reaches **98.75% on the LIBERO VLA benchmark** and implements four MOT trackers from scratch. 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

- **Architected MARS (Mobile Assistive Robotic System)** — a 7-DOF mobile manipulator with hybrid gripper and RGB-D cameras on a **4-layer ROS2 architecture**: WebRTC frontend, Task Manager (LLM planning), hierarchical Skill Controller (bounded retries), Hardware Controllers. Published in *Ubiquitous Robots 2025*.
- Led perception ML for a **production bin-picking system**: improved success from 70% to **92% in cluttered real-world scenes**, processing hundreds of items per hour. In production 3+ years across multiple deployment sites.
- 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.
- Trained and shipped deep learning models for **object detection, segmentation, and 6D pose estimation** on RGB-D data; handled long-tail challenges via active learning and open-vocabulary detection.
- Integrated **VLMs (CLIP, BLIP, LLaVA, GroundingDINO, Gemini Vision)** for zero-shot / open-vocabulary recognition and semantic grounding; implemented **RAG** pipelines over internal knowledge bases.
- Built **LLM-based task planning** with structured outputs grounded in robot skill affordances; multi-agent orchestration with **LangGraph**, episodic + semantic memory (SQLite + ChromaDB).
- Built end-to-end **training pipelines**: dataset curation (synthetic + real + production failures), augmentation, hyperparameter tuning, evaluation harnesses, continuous improvement loops.
- 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 customer deployment teams; mentored 3 junior engineers. Contributed core components to national R&D projects with multi-million USD funding.

Selected Projects

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 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`.

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.
- 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.

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.
- Full-stack platform: FastAPI orchestration + Next.js 14 / Canvas UI with side-by-side comparison and frame-by-frame algorithm stepping; deployed on Docker, Vercel, and RunPod serverless GPU workers.

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.

CareRobotAgent + PyPlanner — LangGraph Multi-Agent Framework *Open Source*

carerobotagent | pyplanner | demo

- **LangGraph** supervisor routing voice (Whisper STT + gTTS), intent, task planning, and robot execution in AI2-THOR with auto-replan on failure; dual memory (SQLite + ChromaDB).
- **PyPlanner**: pluggable library implementing **seven LLM planning methods** (Direct, CoT, Few-Shot, Self-Refine, ReAct, Hierarchical, LLM Router) behind a unified interface, with AI2-THOR-grounded two-layer goal verification and quantitative benchmarking. Backend-agnostic (Ollama / OpenAI / Anthropic).

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

- Novel transformer with a fine-grained attention module for cluttered RGB-D scenes; **20% precision improvement** over CNN baselines on grasp-pose detection.
- Trained on 200K+ annotations with custom augmentation; deployed on NVIDIA Jetson with TensorRT at 30+ FPS.

GroundingDINO Annotation & Training Tool *2024*

github.com/mtbui2010/groundingdino_tool

- Streamlit web tool for semi-automatic labeling and open-vocabulary detector fine-tuning: YOLOv8n → GroundingDINO → FastSAM → human verification → fine-tuning loop.
- Targets the long-tail problem: closes label gaps for novel objects without large pre-labeled datasets.

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

Robotics	mobile manipulation, bin-picking, grasp planning, skill controllers with hierarchical retry, ROS2 / ROS, MoveIt2, hand-eye calibration
Computer Vision	detection, semantic / instance segmentation, 6D pose estimation, multi-object tracking (SORT / DeepSORT / ByteTrack), person re-ID, RGB-D, point clouds, scene understanding
Embodied & Multimodal AI	vision-language-action (OpenVLA-OFT), VLMs (CLIP, BLIP, LLaVA, GroundingDINO, Gemini Vision), LLM planning, RAG, prompt engineering, agent frameworks (LangGraph)
Deep Learning	vision transformers, CNNs, attention, transfer / self-supervised learning, diffusion policies, hyperparameter tuning
Frameworks	PyTorch, TensorFlow, Hugging Face, ONNX, TensorRT, CUDA, LeRobot
Simulation	Isaac Lab / Isaac Sim, MuJoCo, AI2-THOR, LIBERO benchmark
Languages	Python (expert), C++, C, 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, Streamlit, FastAPI, Next.js, Vercel, RunPod
Certifications	Visual Perception for Self-Driving Cars (Coursera, 2023); Improving Deep Neural Networks (DeepLearning.AI, 2021)