Gautham Vinod

+1-(765) 426-9789 · gauthamvinod@outlook.com · linkedin.com/in/gautham-vinod · Google Scholar (199+ citations) · West Lafayette, IN

SUMMARY

PhD candidate (Purdue, Aug 2026) in computer vision and vision-language models. My published work on 3D geometric reasoning has outperformed VLM baselines on peer-reviewed benchmarks. 5+ years of NIH- and NASA-funded research; 12 peer-reviewed publications; 199+ citations; CVPR 2026 Best Paper Runner-up. Seeking AI/ML roles in computer vision, multimodal perception, 3D understanding, or Robotics.

EDUCATION

Ph.D., Electrical & Computer Engineering | Purdue University — Aug 2021 – Aug 2026
Computer Vision, Vision-Language Models, 3D Perception • GPA: 3.84/4.00

M.Sc., Aeronautics & Astronautics | Purdue University — Aug 2019 – May 2021
Dynamics & Control, Deep Learning • GPA: 3.91/4.00

EXPERIENCE

Purdue University — VIPER Lab (NIH-Funded) — Jan 2020 – Present
Lead PhD Researcher, Computer Vision & AI

- **Led the lab's research direction in 3D dietary assessment:** Identified monocular scale recovery as a key open problem in AI-driven precision nutrition, producing 5 publications across CVPR, ICME, CAI, and SSIAI (2022–2026) and an IEEE SSIAI survey now pursued as open problems by 5+ academic groups. Cited 199+ times to date.
- **Translated research systems into production ML pipelines (NIH):** Designed and deployed vision models for real-world patient data analysis for 5000+ participants; built annotation infrastructure processing 15K+ images/month.
- **Real-time egocentric calorie tracking that outperforms frontier VLMs:** Built EgoCal, a stereo-temporal fusion architecture with dual-pathway design analyzing temporal differences and absolute predictions of physical quantities. Outperforms standard vision backbones and state-of-the-art VLMs on the novel EgoEat Dataset, achieving 41.14% PMAE in food weight tracking and 35.12% PMAE in caloric estimation from continuous wearable-camera video. **[CVPRW 2026]** *Runner-up Best Paper, CVPR (MTF) 2026*
- **Beat GPT-5 on volume estimation by 58.2% MAPE on food benchmarks:** Designed a multi-modal framework fusing stereo geometric cues with language priors (CLIP based transformer with text encoders and VLM-inspired projection layers) for object volume estimation without calibration or reference markers. Achieves 22.54% MAPE on MetaFood3D—a 58.2% reduction over GPT-5 (next-best at 53.86%)—and 29.74% MAPE on OmniObject3D. **[CVPRW 2026]**
- **Reduced 3D food volume error by 29.3% MAE from a single image:** Introduced a real-scale 3D reconstruction framework pairing CLIP visual features with multi-view renders of single-view diffusion-based 3D reconstructed models to predict a physical scale factor—addressing the scale-ambiguity limitation in NeRF/Gaussian Splatting pipelines. Outperforms GPT-4o by 29.3% in MAE and 43.3% in MAPE on MetaFood3D; generalizes cross-dataset with 4× improvement over RGB-only baselines. **[CAI 2026]**
- **Architected a first-ever image analysis method for physical quantity change prediction across images:** Developed a VLM-based framework using natural language prompts to localize food items and estimate consumption weight from paired before/after RGB images, eliminating the need for depth sensing, multi-view input, or pixel-level segmentation. Demonstrates consistent improvements over existing approaches across 3 public datasets. **[CVPRW 2026]**

Purdue University — Industry Collaborations — Jan 2020 – Present
Lead AI/ML Engineer

- **Eli Lilly / NIH – Built depth domain adaptation model reducing food energy prediction error by 31%:** Pioneered a depth domain adaptation technique mapping RGB image features into synthetic depth space and energy density space for caloric estimation, enabling generalization across camera types and lighting conditions. **[MIPR 2022]**
- **John Deere – Eliminated 89% of manual decisions in autonomous grain-offloading:** Built a deployable CV pipeline and control system logic on SIMULINK for precise grain unloading from combines. Validated on full-scale John Deere harvester prototypes in field trials. **[IFAC 2022]**
- **NASA ULI – Cut swarm communication bandwidth 73% while holding 94% detection accuracy:** Developed a cooperative multi-agent deep learning framework for search-and-rescue UAV networks with communication-efficient distributed inference.

Blue Wave AI Labs, Indiana — May 2022 – Aug 2022
AI Research Intern — Safety-Critical ML

- Developed uncertainty-quantifying anomaly detection models for nuclear facility sensor telemetry (Random Forest + Gradient Boosting with calibrated confidence estimates). Validated against 2+ years of operational data from 8 US plants; model deployed into production safety monitoring pipeline.

Nanyang Technological University, Singapore — Jan 2019 – May 2019
Research Assistant — Edge AI

- Achieved 22 FPS real-time object detection on ARM processors (Raspberry Pi 4, Jetson Nano) via quantization and model distillation of YOLO v3 and MobileNet-SSD. Integrated into autonomous surface vehicles for surveillance.

SKILLS

Research: Computer Vision, Vision-Language Models (CLIP, ViT, LLaVA, Flamingo), 3D Reconstruction (NeRF, Gaussian Splatting), Generative AI, Diffusion Models, Multi-Modal Learning, Depth Estimation, Contrastive Learning, Large Language Models (LLMs)

Engineering: PyTorch, TensorFlow, Hugging Face Transformers, OpenCV, ONNX, Docker, CUDA, Model Quantization & Pruning

Languages: Python (expert), C++, C#, Java

RESEARCH PROJECTS

Visual-Temporal Fusion for Continuous Object Weight Estimation from Video (Provisional Patent, US App. 64/041,021) Apr 2026

- Co-invented method for continuous, real-time object weight estimation from video via visual-temporal fusion, extending the EgoCal research line into protectable IP. Filed through Purdue Research Foundation with F. Zhu, B. Coburn, and S. Raghavan.

Monocular GPS Tracking from Video — US Navy AI Tracks at Sea Challenge Nov 2020 – Dec 2020

- Top 5% globally: built an end-to-end monocular video pipeline for maritime GPS coordinate prediction achieving 0.099% RMSE, using relative coordinate shifting and object detection pipelines to eliminate reference-frame dependencies. Awarded $3,000 prize by the Naval Research Office.

Electromagnetic Docking Mechanism — Magnetic Landing Gear (Patent IN 201841039774) Feb 2018 – Oct 2018

- Designed an electromagnetic gear achieving 40% faster engagement versus conventional pneumatic systems with reduced fuel and energy use. Patent issued by Controller General of Patents, India.

AWARDS & RECOGNITION

- **CVPR Workshop 2026 (MTF):** Runner-up Best Paper Award **[EgoCal]** Jun 2026
- **US Dept. of the Navy:** AI Tracks at Sea Challenge — Runner-up (top 5%), $3,000 prize Jan 2021
- **Bill & Melinda Gates Foundation:** $100,000 grant for VaxiBead — Smart vaccination records for underprivileged children May 2018
- **NTU India Fellowship:** Research Assistant Fellowship, Nanyang Technological University Jan 2019

PUBLICATIONS

- **G. Vinod**, S. Raghavan*, B. Coburn*, W.-L. Chu, J. Ma, F. Zhu, "EgoCal: Combining Egocentric Data with Vision-Language Fusion for Real-time Dietary Tracking," *IEEE/CVF CVPRW*, pp. 9601–9611, 2026. Stereo-temporal dual-pathway + Kalman fusion; novel EgoEat Dataset; 41.14% PMAE weight, 35.12% PMAE calories — outperforms SOTA VLMs.
- **G. Vinod**, S. Raghavan, B. Coburn, F. Zhu, "DietDelta: A Vision-Language Approach for Dietary Assessment via Before-and-After Images," *IEEE/CVF CVPRW*, 2026. VLM + NL prompt localization for food-item-level weight estimation from paired RGB images; consistent improvements over existing approaches across 3 datasets.
- **G. Vinod***, B. Coburn*, S. Raghavan*, F. Zhu, "Not Your Stereo-Typical Estimator: Combining Vision and Language for Volume Perception," *IEEE/CVF CVPRW*, 2026. First stereo + language fusion model for volume estimation; 22.54% MAPE on MetaFood3D — 58.2% reduction over GPT-5 (next-best).
- **G. Vinod**, F. Zhu, "Food Portion Estimation: From Pixels to Calories," *IEEE SSIAI*, 2026. Survey of 150+ papers; identifies open challenges in depth-agnostic scale perception and cross-cuisine generalization.
- **G. Vinod***, B. Coburn*, S. Raghavan*, J. He, F. Zhu, "Size Matters: Reconstructing Real-Scale 3D Models from Monocular Images for Food Portion Estimation," *IEEE CAI*, 2026. Real-scale 3D reconstruction from single image via CLIP + rendered-view scale factor; 29.3% MAE reduction over GPT-4o on MetaFood3D.
- Y. Chen, J. He, **G. Vinod**, et al., "MetaFood3D: Large 3D Food Object Dataset with Nutrition Values," *arXiv:2409.01996*, 2024. 637 high-accuracy 3D scanned food objects with nutritional ground-truth; benchmark used by 8+ research institutions.
- J. He, Y. Chen, **G. Vinod**, et al., "MetaFood CVPR 2024 Challenge on Physically Informed 3D Food Reconstruction: Methods and Results," *arXiv:2407.09285*, 2024. Technical report synthesizing submissions from 800+ participants across 25 teams; identified algorithmic advances adopted at Google, Meta, and academia.
- J. Ma, X. Zhang, **G. Vinod**, S. Raghavan, J. He, F. Zhu, "MFP3D: Monocular Food Portion Estimation Leveraging 3D Point Clouds," *ICPR Workshops*, Springer, 2024. 3D point cloud projected to 2D for volume regression; 34% accuracy improvement on unseen food categories when fused with monocular pipeline.
- **G. Vinod**, J. He, Z. Shao, F. Zhu, "Food Portion Estimation via 3D Object Scaling," *IEEE/CVF CVPRW*, pp. 3741–3749, 2024. PnP-based pose estimation scales generic 3D food meshes to 2D eating scenes; single-stage inference achieving 8.5% MAPE on real meals.
- Z. Shao, **G. Vinod**, J. He, F. Zhu, "An End-to-End Food Portion Estimation Framework Based on Shape Reconstruction from Monocular Image," *IEEE ICME*, 2023. First unified differentiable pipeline jointly learning 2D segmentation and 3D metric scale prediction in one stage.
- **G. Vinod**, Z. Shao, F. Zhu, "Image Based Food Energy Estimation With Depth Domain Adaptation," *IEEE MIPR*, pp. 262–267, 2022. Depth domain adaptation maps RGB features to synthetic depth space for energy estimation; 31% error reduction with zero-shot cross-camera generalization.
- C. Jiang, S. Dhamankar, Z. Liu, **G. Vinod**, et al., "Co-simulation of the Unreal Engine and MATLAB/Simulink for Automated Grain Offloading," *IFAC-PapersOnLine*, 2022. Modular Unreal Engine/Simulink co-simulation; 89% reduction in human-in-loop decisions; validated on full-scale John Deere harvester.