MINH P. VO

Email: phuocminhvo@gmail.com
Website: https://minhpvo.github.io/

HIGHLIGHTS

15+ years experience in developing and leading Computer Vision and Machine Learning projects at top companies and institutions

Provide mentorship to 50+ interns, engineers, and researchers

Strong records in the academic community (3 awards, 4 patterns, and 30+ publications)

SKILL

Generative image and video synthesis, deep fake detection, multi-modal foundation models, digital human, 4D scene understanding, AI model distillation and deployment.

AWARD

Distinguish Egocentric Papers (CVPR2024)

Best Paper Finalist for Ego4D (CVPR 2022)

Qualcomm Innovation Fellowship (2017)

Measurement Technology and Science Outstanding Paper Award (2014)

WORK EXPERIENCE

SpreeAI   San Francisco, CA
Head of Engineering and Machine Learning   *Jan. 2023 - now*

- Manage and grow a team of researchers and engineers. Increase the team size 6-folds to 25 talented people (largest in the company) since joining.
- Initiate and drive the ML team agenda from research to production for our e-commerce platform with state-of-the-art virtual tryon technology. Workstreams include (1) scalable data processing, (2) tryon image generation, (3) tryon video generation, (4) large vision-language models (VLMs), (5) model distillation for mobile and cloud deployment, (6), automatic quality control, deep fake detection, and (7) adversarial digital watermarking. My team's products enable the company to land product partnerships with Saks Fifth Avenue (strategic partnership), Fanatics, Louis Vuitton, Fendi, Gucci, LVMH, and Adidas.
- Inherit and reboot the mobile, Backend, and DevOps teams. Make them productive again.
- Steer the company roadmap.
- Establish official academic collaborations with CMU, MIT, and UIUC to advance research
- Connect and sponsor tutorials and workshops at top venues like CVPR.

Meta, Reality Labs Research   San Francisco, CA
Research Scientist, Tech Lead   *Sep. 2018 - Dec. 2022*

- Tech lead the Fundamental research team of 5 people for 3D scene perception using Meta'sAria glass. The goal is to create an efficient perception system for Aria glass that can run onboard to detect, track, and forecast scene motion for arbitrary daily activities. Projects include:
  (1) Sensor fusion for 3D object and human detection and tracking. Principally led and co-led EgoHumans, Snipper, BANMo, IDEO, TimeSync (patent pending), and contributed to Lisa, Ego4D, ODAM, ContactOPT. Skill sets include human pose estimation, object pose estimation, sensor fusions, articulated neural radiance, video 6-dof pose estimation, and graph neural network.
  (2) Neural avatar and scene rendering. Principally led the development of TexMesh and ANR (patent pending), and contributed to Open4D and TAVA. Skill sets include illumination understanding, depth

estimation, geometric calibration, human pose estimation, GAN, neural texture and deferred rendering, and neural radiance field.

(3) 3D human action recognition and motion prediction with scene context. Principally led the development of HMP and STSS. Skill sets include human pose estimation, depth estimation, motion prediction, visual saliency, and spatiotemporal sparse processing.

- Collaborate with teams in Meta and academic partners. Point of contact for research collaboration with the Robotics Institute, Carnegie Mellon University.

The Robotics Institute, Carnegie Mellon University   Pittsburgh, PA
Research Associate - Mentors: Srinivasa Narasimhan and Yaser Sheikh   Oct. 2012 - Aug. 2019

- 4D event browsing: Novel method for image-based browsing of dynamic event from multiple views.
- Self-supervised scene adaptive human appearance descriptor: Develop a novel framework to associate and track multiple people in highly chaotic scenes.
- ShapeFusion: Develop a novel generic framework for accurate 3D tracking and structured keypoint detection for rigid objects such as car.
- Spatiotemporal calibration for dynamic 3D reconstruction: Develop a novel spatiotemporal bundle adjustment algorithm for multiple uncalibrated and unsynchronized smart phone videos in the wild.
- Structured light on highly textured object: Develop a novel texture-illumination separation algorithm enabling single-shot structured light systems to produce dense 3D shape of highly textured objects.
- Passive tomography of turbulence strength: Develop a novel and inexpensive method to estimate the turbulence strength using multiple off-the shelf-cameras.
- Panoptic studio: Develop an accurate and automatic geometric calibration algorithm for a virtualization studio consisting of more than 500 cameras and 6 projectors.

Adobe Research   San Jose, CA
Research Intern - Mentors: Sunil Hadap, Kalyan Sunkavalli, Ersin Yumer   May. 2017 - Aug. 2017

- Spatiotemporal human tracking from multiple video cameras in the wild.

Microsoft Research   Redmond, WA
Research Intern - Mentors: Neel Joshi and Sudipta Sinha   May. 2016 - Aug. 2016

- Direct simultaneous camera calibration and depth estimation for small baseline videos.

Mechanical Dept., Catholic University of America   Washington, DC
Research Associate - Mentor: Zhaoyang Wang   Nov. 2009 - Aug. 2012

- Structure light calibration: Develop an accurate, fast, and flexible geometric calibration approach. Real time measurement of 0.005% relative accuracy a 10fps was achieved.
- 2D-3D Digital Image Correlation (DIC): Develop an accurate image matching algorithm for strain/stress measurement of deforming structure. Our synthetic tests reveal that the developed algorithm can estimate the particle displacement at 5000 point/sec with relative accuracy better 0.001%.
- Camera calibration: Develop an inexpensive and accurate geometric camera calibration algorithm. The calibration error is several times smaller than other widely-used packages (i.e., OpenCV, Caltech Calib).

PATENT

1. **M. Vo**, "Virtual Garment Fitting Using Machine Learning,", US Patent (Pending)
2. **M. Vo**, C. Lassner, A. Raj, C. Stoll, "Articulated Deferred Neural Rendering for Photorealistic Human Avatars from Videos,", US Patent 11,651,540
3. **M. Vo**, K. Somasundaram, S. Lovegrove, "Time-synchronized Distributed Passive Captures,", US Patent 11,762,080

4. N.S Joshi, S.N Sinha, **M. Vo**, "Scene reconstruction from bursts of image data," US Patent 10,535,156

PUBLICATION

Refereed Journal Publications

1. **M. Vo**, K. Sunkavalli, E. Yumer, S. Hadap, Y. Sheikh, S.G. Narasimhan, 'Self-supervised Multi-view Person Association and Its Applications," IEEE Trans. PAMI 2020.

2. **M. Vo**, Y. Sheikh, S.G. Narasimhan, 'Spatiotemporal Bundle Adjustment for Dynamic 3D Human Reconstruction in the Wild," IEEE Trans. PAMI 2020.

3. **M. Vo**, S. G. Narasimhan, and Y. Sheikh, "Texture illumination separation for single shot structured light reconstruction," IEEE Trans. PAMI 2015

4. Z. Wang, **M. Vo**, H. Kieu, T. Pan, "Automated Fast Initial Guess in Digital Image Correlation," Strain, 2014.

5. H. Kieu, T. Pan, Z. Wang, M. Le, H. Nguyen, **M. Vo**, "Accurate 3D shape measurement of multiple separate objects with stereo vision," Measurement Science and Technology, 2014.

6. T. Nguyen, H. Nguyen, **M. Vo**, Z. Wang, L. Luu, and J. Ramella-Roman, "Three-dimensional phantoms for curvature correction in spatial frequency domain imaging," Biomedical Optics Express, 2012.

7. **M. Vo**, Z. Wang, B. Pan, and T. Pan, "Hyper-accurate flexible calibration technique for fringe-projection-based three-dimensional imaging," Optics Express, 2012.

8. **M. Vo**, Z. Wang, L. Luu, and J. Ma, "Advanced geometric camera calibration for machine vision, Optical Engineering, 2011.

9. L. Luu, Z. Wang, **M. Vo**, T. Hoang, and J. Ma, "Accuracy enhancement of digital image correlation with B-spline interpolation," Optics Letters, 2011.

10. T. Hoang, Z. Wang, **M. Vo**, J. Ma, L. Luu and B. Pan, "Phase extraction from optical interferograms in presence of intensity nonlinearity and arbitrary phase shifts," Applied Physics Letters, 2011.

11. **M. Vo**, Z. Wang, T. Hoang, and D. Nguyen,"Flexible calibration technique for fringe-projection-based three-dimensional imaging," Optics Letters, 2010.

Refereed Conference Publications

1. X. Zhao, C. Gao , X. Rong, S. Saito , **M. Vo**, J. B. Huang, A. Schwing , "On the Non-rigid Motion for Modeling Animatable Avatars from Monocular Videos", in submission

2. R. Khirodkar, A. Bansal, K. Kitani, **M. Vo** , "Ego-Humans: An Ego-Centric 3D Multi-Human Benchmark", ICCV 2023 Oral Presentation

3. T. Do, L. Lemke, J. Guo, K. Vuong, **M. Vo**, H.S. Park, "IDEO: Large Scale Egocentric 3D Object Dataset and Benchmark Challenges", in submission

4. S. Zou, Y. Xu, C. Li, L. Ma, L. Cheng, C. Sweeney, R. Newcombe, **M. Vo**, "Snipper: A Unified 3D Transformer for Simultaneous Multi-person 3D Pose Estimation Tracking and Forecasting on a Video Snippet," IEEE Trans. on Circuits and Systems for Video Technology, 2023.

5. K.M. Nguyen, M. Do, K. Somasundaram, **M. Vo**, "Self-Attention Based Spatiotemporal Sampling for Egocentric Action Recognition in Map-Grounded Videos," IEEE Trans. Image Processing, 2023

6. R. Li, J. Tanke, **M. Vo**, M. Zollhoefer, J. Gall, A. Kanazawa, C. Lassner, "TAVA: Template-free Animatable Volumetric Actors", ECCV 2022.

7. G. Yang, **M. Vo**, N. Neverova, D. Ramanan, A. Vedaldi, H. Joo , "BANMo: Building Animatable 3D Neural Models from Many Casual Videos," CVPR2022 Oral Presentation.

8. E. Corona, T. Hodan, **M. Vo**, C. Sweeney, R. Newcombe, F. Moreno, L. Ma, "LISA: Learning Implicit Shape and Appearance of Hands," CVPR2022.

9. K. Grauman et. al, "Ego4D: Around the World in 3,000 Hours of Egocentric Video," CVPR, 2022 Oral Presentation. Best paper finalist

10. K. Li, D. DeTone, S. Chen, **M. Vo**, J. Straub, "Frodo++: Realtime Monocular 3D Object Spatialization by Relational Reasoning," ICCV, 2021 Oral Presentation

11. P. Grady, C. Kemp, C. Tang, C. Twigg, **M. Vo**, S. Brahmbhatt. "ContactOpt: Optimizing Contact to Improve Grasps," CVPR, 2021 Oral Presentation

12. A. Raj, J. Tanke, **M. Vo**, C. Stoll, C. Lassner, "ANR: Articulated Neural Rendering for Virtual Avatars," CVPR, 2021.

13. Z. Cao, H. Gao, K. Mangalam, Q.Z. Cai, **M. Vo**, J. Malik, "Long-term Human Motion Prediction with Scene Context," ECCV, 2020 Oral Presentation

14. T. Zhi, C. Larssner, T. Tung, C. Stoll, S.G. Narasimhan, **M. Vo**, "TexMesh: Reconstructing Detailed Human Texture and Geometry from Monocular Video," ECCV, 2020.

15. A.Bansal, **M. Vo**, Y. Sheikh, D. Ramanan, S.G. Narasimhan, "4D Visualization of Dynamic Events from Unconstrained Multi-View Videos ," CVPR, 2020.

16. D. Reddy, **M. Vo**, S.G. Narasimhan, "Occlusion-Net: 2D/3D occluded keypoint localization using graph networks ," CVPR, 2019.

17. D. Reddy, **M. Vo**, S.G. Narasimhan, "CarFusion: Combining Point Tracking and Part Detection for Dynamic 3D Reconstruction of Vehicles," CVPR, 2018.

18. **M. Vo**, S.G. Narasimhan, Y. Sheikh, "Spatiotemporal Bundle Adjustment for Dynamic 3D Reconstruction," CVPR, 2016.

19. M. Alterman, Y.Y.Schechner, **M. Vo**, S. G. Narasimhan, "Passive tomography of turbulence strength," ECCV, 2014.

20. **M. Vo**, S. G. Narasimhan, and Y. Sheikh, "Separating Texture and Illumination for Single-Shot Structured Light Reconstruction," CVPRW, 2014.

EDUCATION

Carnegie Mellon University, Pittsburgh, PA
Ph.D. in Robotics
Advisors: Srinivasa Narasimhan and Yaser Sheikh
Thesis: Exploiting Point Motion, Shape Deformation, and Semantic Priors for Dynamic 3D Reconstruction in the Wild.

Carnegie Mellon University, Pittsburgh, PA
M.Sc. in Robotics
Advisors: Srinivasa Narasimhan and Yaser Sheikh
Thesis: Texture and Illumination Separation for Single-shot Structured Light Reconstruction.

Catholic University of America, Washington, D.C
B.E. in Electrical Engineering
Summa Cum Laude
Advisor: Zhaoyang Wang
Thesis: High accuracy camera calibration and its application.

SELECTED COURSEWORK

Computer Vision, Geometry-based Vision, Learning-based Vision, Compressive Sensing and Sparse Optimization, Statistical Methods for Robotics, Machine Learning.

RESEARCH INTEREST

Computer Vision 3D Vision, Scene Reconstruction, Virtual Human, and Action Understanding.
Computational Photography Video Super-resolution, View Synthesis.
Machine Learning Semi-supervised Learning, Structured Learning.

PROFESSIONAL ACTIVITY

Reviewer: CVPR, ICCV, ECCV, TPAMI, TIP, SIGGRAPH.