Statement of Interest

Yifei Dong
ETH Zurich

Every day, I can feel the world I live in being driven by computer vision. Within the realm of autonomous driving, robotics and medical science, computer vision often serves as the most effective tool, stealing the spotlight throughout the world. As a passionate graduate student who loves computer vision, I am always intrigued by the efficiency of practical vision techniques and the charm of cutting-edge research. I, therefore, would like to apply to the R&D position with a specialization in computer vision and deep learning.

Inspired by my tutor at ETH Zurich, Prof. Roland Siegwart, my focus in the master's program is perceptive robot manipulation. The reason for choosing it lies in that, in both industrial and social domains, autonomous systems are expected to be highly interactive with their surroundings, and thus machine perception is one of the most important keys of achieving it. With the profound progress in deep learning, semantic, photometric and geometric information is more effectively interpreted from visual input. The development in depth camera and stereo vision are demystifying the magic of perception, and closing the gap between us and artificial intelligence. After my past experiences of master thesis and projects in computer vision, I feel more and more intrigued by the science of computer vision and am determined to dedicate my passion to this field.

The rigorous curriculum in ETH Zurich has built a solid academic foundation for me. To consolidate my basis of knowledge, I took courses in computer vision, optimization, estimation and deep learning. In several courses on 2D and 3D computer vision, I investigated the state-of-the-art image processing techniques, as well as the learning-based detection and segmentation approaches, and took the lead of a team project on 3D reconstruction. I also audited deep learning related courses at ETH, as well as the course of Deep Learning in Coursera, which gives me great insight into implicit visual information extraction from deep neural networks. I learnt how to leverage Kalman filters, particle filters, Bayesian inference, etc, to recursively estimate the object's pose through prediction and innovation. In Dynamic Programming and Optimal Control, as well as Model Predictive Control, I grasped general approaches that are ubiquitously applied in solving numerical constrained optimization problems in computer vision and robot control, for instance, SLAM and deep learning.

To enhance my understanding of SLAM and stereo vision, I took part in a team project in collaboration with Mixed Reality & AI Lab, Microsoft, supervised by Prof. Marc Pollefeys with a topic of object mesh reconstruction using RGB-D cameras. We propose a novel pipeline of the extraction of objects of interest from several dense point clouds of indoor scenes, and generate smooth mesh for the objects. On the other hand, we propose a heuristic fitting and deep learning approach to rectify depth images and, in this way, the depth offsets are compensated for and the ToF camera accuracy increases considerably.

Computer vision is vital to robots, and we could say, without it the robots would be blind. In my semester project at Robotic Systems Lab, ETH Zurich, with a topic of model predictive control for door manipulation, I innovatively proposed a legged robot system capable of multi-contact multi-object manipulation, such that the leg of a legged robot is thus able to switch swiftly from trotting to pushing a door. However, a perception module is required in the object-augmented whole body planner of the robot. Otherwise, the robot could basically do nothing with doors that it is not familiar with. With a focus on finding out a solution to that, my master thesis at RSL on perceptive mobile estimation of door's pose and dimension, supervised by Prof. Marco Hutter, complements my previous research. In the thesis, I developed one of the first reliable frameworks of door state estimation and parameter identification for mobile manipulation. I trained a customized neural network for door and handle detection based on YOLOv5. Image augmentation and hyperparameter evolution are introduced, which enormously improve the prediction precision and effectively suppress the influence of ambient light, visual obstacles and motion blur. By fusing RGB-D images and robot odometry, a novel 3D door tracking algorithm is presented. The work is validated in a hardware experiment on a legged robot for visual servoing handle grasping. A dataset on various doors is created, composed of RGB-D data, robot odometry and ground truth.

Besides research programs in 3D vision, I have good knowledge of traditional image processing approaches as well. Back in my undergraduate studies in Shanghai, I was on the FSAE autonomous racing car team. I am in charge of the development of autonomous lane keeping assist systems and traffic sign recognition. The research contributed to the whole driving assistance team and won us excellence award in BAIC automobile invitational tournament in Beijing.

To obtain more insights and experience in industrial computer vision application, I had two internships in the autonomous driving departments of NIO Inc. and SAIC Motor, respectively. In NIO, I implemented LiDAR simulation of autonomous vehicles in various traffic scenes to collect as much information as possible of vehicles and pedestrians in the neighborhood. Recently, I am applying my expertise in computer vision to an industrial service robot team at F&P Robotics in Zurich. My focus is on high-level skills (door opening, drink distribution) design and development on the platform of a mobile service robot. I am developing a real-time handle detector combining gripper camera and radar information in a neural network. It provides me with an industrial perspective of vision systems development, and clarifies my goal of deeper research in the field of computer vision.

Last but not least, I have started a short part-time contract with Vitestro, a medical robotics company providing autonomous blood drawing solutions. My main tasks include processing image data from a 2D ultrasound probe, reconstructing vein segments, and finding a suitable puncture location. It offers me a great opportunity to know more about computer vision applications in medical and surgical fields.

Thanks to the research and industrial experiences above, I have cultivated the abilities of independently coming up with new ideas from state-of-the-art research. In the meantime, I am also continuously improving my object-oriented programming skills of C++ and Python, as well as the proficiency in computer vision algorithms in daily Leetcode practices. In a nutshell, I have lit a candle of curiosity of science and passion for computer vision. I am just waiting for an opportunity to lighten the world. I look forward to the R&D position and to bringing my computer vision expertise gained from past experience to the future broader stage.