Research Statement

My research interests are computer graphics, computer vision, and deep learning, especially focus mostly on depth map processing, motion estimation with scene flow and exploring applications of artificial intelligence. The central motivating theme of my research is moving more intelligence to end devices: the use of latest technologies transferred from laboratory to make device more comprehensive, reasonable and perceivable.

Due to the restriction of CMOS imaging techniques and the limitation of algorithms, there is an enormous gap between machines' ability to understand images and comprehension of humans. In my research, I fill some of these gaps by reconstructing high frame rate depth maps, estimating 3D motion from color images and depth maps (RGB-D) and applying latest deep learning algorithm in practical application.

My master's and doctoral research have provided me with strong expertise in various fields, such as computer graphics, computer vision, deep learning, computer architecture, and embedded micro-systems. I will leverage my skills to create essential techniques of motion estimation and image processing benefiting for motion capture and mechanical control, and to make on-device (e.g. mobile phones, surveillance cameras) comprehensions ubiquitous with efficient AI hardware (e.g. Google TPU, Intel movidius) and algorithmic advancements.

Current Work

My current research centers on temporal upsampling of depth maps to make up for the low frame rate of TOF depth camera. I proposed a method to reconstruct the high frame rate depth maps with the consecutive images captured by a hybrid camera (high frame rate color images and low frame rate depth maps). In this method, various techniques including optical flow, scene flow, geometry registration, and GPU programming were adopted to improve the low frame rate depth cameras. In order to evaluate the performance of reconstruction, a RGB-D dataset with over 20,000 consecutive images, captured by our hybrid camera hardware, was especially built in this work. And a GPU-based numerical solution was exploited to reconstruct more efficiently. Through the experiments on the proposed dataset and the publicly available dataset, our method outperforms the state-of-the-art methods and contributes to the fields of human motion capture, image processing and video editing. This work was published on IEEE Transactions on Visualization and Computer Graphics (TVCG, IF=2.84) [J1]. It is the data acquisition method of the research "3D Human Motion Modeling Technology", whose purpose is to capture the 3D geometric model and joint point of human body, won the first prize of Science and Technology Progress awarded by the China Computer Federation (CCF, the largest computer association in P.R. china) in 2018.

In the course of the study mentioned above, I found the accuracy of reconstructed depth map is reduced by inevitable flaws in the images captured by the hybrid camera, such as color images with motion blur, depth maps with noise. To eliminate the flaws and get perfect images, I tried to upgrade hybrid camera hardware by research the color and depth camera imaging sensors and technologies (e.g. CCD, CMOS, TOF, structured-light). At the same time, the result of evaluating the development boards of TI OPT8241 depth camera and Sony IMX 219 color camera prove that motion blurring is caused by insufficient light and noise in depth maps is caused by too strong light in the environment. The most feasible way of capturing data is using the color camera with a larger pixels size of sensor and avoiding strong light. These experience guided an adviser's application for the National Natural Science Foundation of China (NSFC) project for manufacturing the hybrid camera.

Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS)
yuanmingze2014@gmail.com • *yuanmingze.github.io*

With those higher frame rate and lower noise images, I further estimated more accurate motion information with scene flow method. We proposed a novel learning based framework to estimate scene flow, which takes both brightness and scene flow losses. Given a pair of RGB-D images, the brightness loss is used to measure the disparity between the first RGB-D image and the deformed second RGB-D image using the scene flow, and the scene flow loss is used to learn from the ground truth of scene flow. We built a convolutional neural network to simultaneously optimize both losses. Extensive experiments on both synthetic and real-world datasets show that our method is significantly faster than existing methods and outperforms state-of-the-art real-time methods in accuracy. This work was concluded and published on BMVC 2018 [C1].

The above is the content of theoretical research. In order to make real-world end device smart enough to recognize motion, object, etc., I tried to address this challenge with latest AI technologies. For this reason, I spent the last year of Ph.D. to carry out the transformation of research achievements, to develop AI based personal protective equipment in safety, especially in the field of working with potentially dangerous voltages. The team composed of 5 interns supervised by me developed a device to verify the worker's identity and the wearing of protective tools from the images captured by the surveillance camera with the detection and classification methods based on CNN. To prevent workers from dangerous and erroneous operations, we developed a system using LSTM based method to identify workers' behavior with acceleration information captured by LPMS-B2 sensor wearing on the wrist. Currently the results of application in the State Grid Corporation of China which is the largest utility company in the world illustrate the AI method can effectively protect worker's safety. However, I found the GPU for CNN's inference is unaffordable for most application scenario, but AI hardware devise new way. This development experience greatly encouraged me to do in-depth research in computer vision, edge computing, and artificial intelligence.

Future Work

My short-term research goal is to more accurately estimate the motion information from image and reconstruct high frame rate depth map with the CNN based method.

The goal of the next five years is making on-device AI ubiquitous. I will take advantage of my knowledge on computer system structure, computer vision, deep learning and hardware knowledge, and make use of artificial intelligence hardware (e.g. TPU, Movidius) and Edge Computing technology to improve the intelligence of devices such as surveillance cameras, mobile phones, etc.