Video reconstruction from undersampled measurements

Prasan Shedligeri[1]

[1]Department of EE, IIT Madras, Chennai, ee16d409@ee.iitm.ac.in

The usual paradigm of *computational photography* is to jointly design hardware and software for specific applications to enhance overall performance and efficiency. However, more often than not, devising new hardware for each application is cumbersome and expensive. These hardware setups are often bulky and require meticulous calibration, making them ill-suited for practical, real-world scenarios. My research tackles the well-established computational photography problems and achieves better performance by re-purposing existing hardware with minimal changes. While this seems ambitious, my research uses state-of-the-art deep learning techniques to achieve this goal.

While deep-learning algorithms achieve significantly higher performance on several tasks, they still require informative inputs. Images and videos captured using traditional cameras have their limitations and are not always suitable for several applications. For instance, a learning-based algorithm proposed for video reconstruction from a single blurred image suffers from motion ambiguity leading to poor reconstructions. An additional input that encodes the motion information would significantly improve the algorithm performance. Hence, in my research, I go beyond the traditional cameras and explore novel sensors and setups most suitable for my application. Along with the suitability, aspects like the hardware's power consumption, commercial availability, form factor are also considered. After choosing the proper hardware, the task is to design a novel learning-based algorithm for each application. My research considers two well-studied computational photography problems: high-frame-rate video and light-field video reconstruction.

I chose two novel sensors for high-frame-rate video reconstruction: neuromorphic event sensors and multi-bucket coded exposure sensors. While event sensors acquire the motion information as discrete, binary events, coded exposure sensors temporally multiplex a video into a single frame. Devising efficient ways to reconstruct the videos from these sensors required the knowledge of state-of-the-art deep learning techniques and an understanding of the sensor mechanism. For instance, I had to use recurrent networks to encode motion information from event sensors and employ shift-varying convolutional layers for processing coded-exposure images. I have also embraced a multi-disciplinary and a curiosity-driven approach in my research, which led me to learn about computational displays in my early years. This knowledge about displays proved helpful in my last year when I got stuck solving the challenging problem of light-field video reconstruction. I also had to develop a fully self-supervised technique to overcome a lack of high-quality ground truth training data for light-field videos. This project gave me an insight into the world of learning from limited data and the new wave of self-supervised approaches in computer vision.

I have obtained ample experience in publishing peer-review papers in top conferences and journals. I have also written several grants securing two travel scholarships worth USD 6000 and was a finalist in the prestigious Qualcomm Innovation Fellowship for 2018. I intend to continue working on light-field video reconstruction for mobile devices using monocular videos and dual-pixel sensor videos. I am always excited to collaborate and work on new areas involving images, videos, and deep learning.