

Pixels at Scale: High Performance Computer Graphics and Vision

Session co-chairs: John Owens, University of California, Davis, and David Luebke, NVIDIA Research

The smartphones we carry in our pockets have remarkable capabilities that were unimaginable only a decade ago: a high-quality retinal display powered by high-performance graphics hardware, a high-resolution camera capable of capturing a billion pixels every second, and a high-bandwidth connection to a cloud infrastructure with tremendous computational horsepower. In short, what we have today is an abundance of pixels that we can produce, process, and consume easily and cheaply. Our session will address the question, “what do we do with all these pixels?”.

The ascendance of the pixel is the culmination of numerous technical advances that began with the invention of computer graphics in the 1960s and digital photography in the 1970s. Modern consumer hardware has made ubiquitous:

- powerful computer graphics hardware, continuously increasing performance and quality of graphics;
- high-resolution displays, approaching the native resolution of the eye; and
- high-resolution low-cost digital cameras, generating trillions of digital photos for analysis and training.

These advances give us traction on two long-standing grand challenges that center on the pixel: interactive, immersive, photorealistic computer graphics and ubiquitous, robust image analysis and understanding. We will discuss four interlocking technology and application areas, spanning both “pixels in” and “pixels out”: computer vision and image understanding, modern computer graphics hardware, computational display, and virtual reality.

These advances give us traction on two long-standing grand challenges that center on the pixel: interactive, immersive, photorealistic computer graphics and ubiquitous, robust image analysis and understanding. We will discuss four interlocking technology and application areas, working roughly from “pixels in” to “pixels out”: computer vision and image understanding, modern computer graphics hardware, computational display, and virtual reality.

Fueled in part by the deluge of pixels – the availability of images for training at massive scale – advances in machine learning are bringing about a sort of Golden Age of computer vision. Many of the traditional grand challenges in machine learning, such as outperforming humans at recognizing objects or understanding speech, have fallen. Computer vision researchers can now tackle problems and applications of image understanding that were previously hard to imagine.

Computer graphics hardware is the computational substrate for the pixel revolution. The graphics processing unit (GPU) in today’s PCs and smartphones represent decades of co-evolution between graphics algorithms and the silicon architectures that execute them. In the process the modern GPU has grown from a fixed-function co-processor to general-purpose

parallel computing platform, in the process accruing considerably more computational horsepower than the rest of the processors in the device put together. The GPUs on which consumers play video games execute tens of thousands of concurrent threads, providing a level of massively parallel computation that was once the exclusive preserve of supercomputers. Thus today's GPUs not only render video games, they also accelerate astrophysics, video transcoding, image processing, protein folding, seismic exploration, computational finance, heart surgery, self-driving cars – the list goes on and on. Importantly, machine learning algorithms (particularly convolutional neural nets or “deep learning”) map especially well to GPUs, which largely power the computer vision renaissance.

Pixels are just data until a display turns them into photons, and display technology is undergoing its own tectonic shifts. LCD and OLED panels are following their own Moore's Law and achieving breathtaking advances in resolution, cost, size (both large and small), brightness – almost every metric one can think of. Less obvious is a body of work in *computational display*, which co-designs the optics and electronics of the display system with the rendering algorithms that generate the pixels. For example, stacking multiple panels can create a light field display, providing glasses-free “3D” (stereo) views with correct motion parallax as the viewer moves. Other novel optics coupled with rendering algorithms enable new tradeoffs, such as trading resolution for a thinner display, or focus cues.

Right now, perhaps the most talked-about applications in the ongoing pixel revolution are virtual and augmented reality. We are captivated by the concept of rendering a virtual world so effectively that the perceptual system accepts it as reality, or by the prospect of seamlessly integrating synthetic information and objects into our view of the real world. Virtual and augmented reality pose huge challenges for all the topics discussed above: computer vision must track the user's slightest motions and gestures, and interpret their environment; graphics hardware must render at unprecedented levels of performance to achieve immersion; and displays must evolve from today's boxy headmounts to something as vanishingly unobtrusive as a pair of eyeglasses.

Our session will begin with Dr. Gordon Wetzstein of Stanford University. Prof. Wetzstein has pioneered the relatively young field of computational display, working on the boundary of optics, electronics, and computer graphics to design innovative display systems with entirely new capabilities. Next, we welcome Dr. Warren Hunt of Oculus. Oculus, now owned by Facebook, is pioneering virtual reality headsets. Dr. Hunt's (and Oculus's) work is fascinating for two reasons: (1) the traditional assumptions in computer graphics are upended when the display is an inch from your eye and controlled by your head; (2) immersive virtual reality requires performance and resolution significantly beyond what current systems offer. We will then hear from Dr. Kristen Grauman, associate professor of computer science at the University of Texas. Prof. Grauman is an expert in computer vision, with particular expertise in the interface between vision and machine learning. Her research couples image recognition with learning from that recognition, with applications in visual search, and a recent focus on first-person vision (enabled, in turn, by advances in cameras) where the camera wearer is an active participant in visual observation. Finally we will conclude with Dr. Kayvon Fatahalian, assistant professor of

computer science at Carnegie Mellon University. Prof. Fatahalian's research couples a systems mindset with deep expertise in pixel processing hardware and software. He will discuss the challenges and opportunities of processing live pixel streams on vast scales, with applications ranging from personal to urban to societal.