

Case ID:M22-111P^  
Published: 1/11/2023

## Multi-Resolution Visual Sensing Architecture for Augmented Reality

-Through the lens of their cameras, mobile devices can visually observe a user's environment for detecting faces, understanding spatial scene geometry, and capturing images and videos. This has yielded a wide range of benefits, especially as image sensors have grown to support increasingly higher resolutions and frame rates. With such precision, it is now possible to position augmented reality (AR) overlays over users' faces or over spatial living environments for entertainment. Virtual AR media can also annotate physical environmental surfaces, including for navigational guidance (e.g., for walking directions).

Unfortunately, visual systems on mobile systems are limited in their spatial precision, computational performance, and energy efficiency while performing continuous visual tasks. Mobile systems are constrained by their small form factors with limited battery sizes and heat management requirements. To reduce power consumption, current systems capture and process entire image frame streams at uniform (reduced) spatial resolutions and uniform (reduced) frame rates, especially by reducing the memory traffic of the DRAM-based frame buffers. However, most scenes do not have the same resolution needs across the entire image frame. For example, precise AR placement requires high resolution for visual features on tracked surfaces, but would suffice with a relatively lower resolution for the rest of the frame. Or, the detection and tracking of faces, hands, and objects could use a higher resolution to capture quick motions, while the rest of a relatively static scene would suffice with a lower frame rate. Hence, there is a need for a visual computing system that captures and processes entire image frame streams at non-uniform spatial resolutions and non-uniform frame rates.

Researchers at Arizona State University (ASU) have developed a visual computing system that enables region-based resolution control. The system allows application developers to dynamically adapt the spatial resolution and update rate of different "rhythmic pixel regions" in a scene. The system ingests pixel streams from image sensors and only encodes relevant pixels prior to storing them in memory. Streaming hardware decodes the stored rhythmic pixel region stream into traditional frame-based representations to feed into standard computer vision algorithms. Additionally, developers can flexibly specify region labels. System has been evaluated on a FPGA platform over a variety of vision workloads and shows significant reduction in interface traffic and memory footprint while providing controllable task accuracy.

Related publication: [Rhythmic Pixel Regions: Multi-resolution Visual Sensing System towards High-Precision Visual Computing at Low Power](#)

Potential Applications:

### Inventors

Venkatesh Kodukula  
Alexander Shearer  
Van Nguyen  
Srinivas Lingutla  
Robert LiKamWa

### Contact

Physical Sciences Team

- ASU's visual computing system can be incorporated into:
  - Video pipelines of commercial systems on a chip (SoCs) used for gaming, mobile devices, augmented reality (AR), virtual reality (VR), etc.
  - Operating system service (e.g., camera APIs)

Benefits and Advantages:

- Different parts of a scene are captured at different spatiotemporal resolutions
- Rhythmic pixel encoder and decoder – decimates the incoming pixel stream while writing to memory and reconstructs the pixel stream on-the-fly while reading from memory
- Demonstrated significant reduction in memory traffic (i.e., minimal DRAM traffic) with controllable accuracy loss