A Visual Cloud for Virtual Reality Applications

Magdalena Balazinska, Luis Ceze, Alvin Cheung, Brian Curless, and Steve Seitz
University of Washington

Our ability to collect images of the world en masse can revolutionize how we interact with the world by enabling powerful virtual reality (VR) applications in education, tourism, gaming, and others. These new applications, however, require a dramatic improvement in technology to manage the massive-scale visual and audio data necessary for truly immersive experiences. In the Computer Science & Engineering Department at the University of Washington, we are building a Visual Cloud that will provide seamless access to a new stack of hardware and software components that, together, will enable the efficient management of massive-scale image and audio data and VR applications.

Motivating Application

Our first target VR application is one in the spirit of Google Streetview, where users will be able to freely move through a virtual rendering of various indoor and outdoor spaces. Example experiences that we seek to provide include (1) an immersive experience for selected outdoor sites (*e.g.*, Seattle's Space Needle) (2) immersive tour of selected places such as the university campus or walkthroughs of museums, and (3) immersive experience of dynamic places such as Seattle's Pike Place Market.

One of the most compelling technologies for VR are light fields. When a person looks in a particular direction or takes a photo, she is observing a subset of the light in the scene; in particular, the subset that hits the retina or camera sensor. If one could capture and represent all of the light in the scene that flows along all rays in space, one could simulate any view of that scene "simply" by selecting the rays that pass through that viewpoint.

Light fields naturally take the form of multidimensional arrays as the set of light rays is six dimensional (an RGB function over a 6D domain): 3D for position, 2D for direction, and 1D for time. Fortunately, there are methods to bring down the dimensionality a bit; modest restriction of the navigable volume reduces it to 5D, and working with a single time instant for static spaces brings it down another dimension to 4D. Further, we can exploit the fact that a user accesses only a 2D "slice" of this data (an image or pair of images) at any given time.

A Visual Cloud

The ability to capture, store, transmit, and render light fields to VR headsets represents a major breakthrough, as it would allow people to experience a remote scene as if they were physically there. Captured images or videos from many viewpoints would represent the raw light field, along with multi-channel/directional audio. This raw data would then be processed and stored before transmission to a viewer, where it would be rendered to the viewpoint of each eye in the tracked VR headset and the audio rendered to each ear, allowing the viewer to walk around the scene and have a feeling of immersion and teleportation to another place. Supporting this at scale is beyond today's computer systems. Our end goal is to develop a full system stack, running in a public cloud, for the ingest, storage, processing, and streaming of visual and audio data in support of VR applications.

Challenges

Building a visual cloud raises many interesting challenges:

- **High-Rate Data Ingest:** To provide immersive experiences, we must first collect large volumes of audio and video data from the scenes of interest. This process can easily generate megabytes or gigabytes of raw data per second. The first challenge is thus to efficiently store the resulting large-bandwidth data streams. One approach is to compress the data at the source and to leverage the elasticity of the cloud to parallelize the data ingest.
- **Data Processing:** The transformation of the raw input data into light-field array is an expensive process. The cloud is well suited to provide the required compute resources and one interesting approach is to employ the most appropriate compute substrate (CPU, GPU, SIMD, FPGA) for various components of the algorithms to maximize performance.
- **Light Field Data Storage:** The output from the offline processing are light field arrays. Given the multidimensional nature of this data, it makes sense to adopt a multidimensional array storage model. However, new data layout and compression algorithms can improve performance. For example, as more image data becomes available, initially sparse arrays become increasingly dense and the storage system must handle this evolution efficiently. Additionally, light fields are naturally represented at different resolutions, e.g., for far-away vs nearby objects. The storage subsystem can take advantage of this property to lay out the data in a way that enables fast retrieval of different resolutions.
- **Distributed Runtime & Scheduling:** To enable realtime VR applications, a final challenge is to build a runtime system and scheduler that ensure efficiency for offline pre-computations and low-latency for real-time applications. The runtime can employ approximate computing and predictive models to selectively stream data to the user device and push computation to the edges to minimize latency.