Bringing greater reality to virtual reality

By Andrew Myers

As impressive and lifelike as virtual reality (VR) systems are today, they don't quite produce the ultra-high-resolution video necessary for optimal "immersive" experiences … yet. Exactly why that is has a lot to do with bandwidth—all that high resolution imagery and rapid framerates chew up data, fast. The computers can't keep up. In a new paper, researchers at Stanford University recently explored a clever solution to the problem by mimicking the eye's own mechanisms for sight.

The researchers built an innovative prototype of a "foveated" compression system that works in conjunction with gaze tracking. The fovea is the ophthalmological term for the area of the retina with greatest visual acuity.

This combination of eye-tracking, compression, and ultra-low latency is, the team believes, a first in Internet-based VR video systems. It is based on a client-and-server model, in which the client is the virtual reality headset. The client tracks and reports the viewer's current gaze position for every frame, allowing the server to encode the next frame foveated on the viewer's precise gaze position.

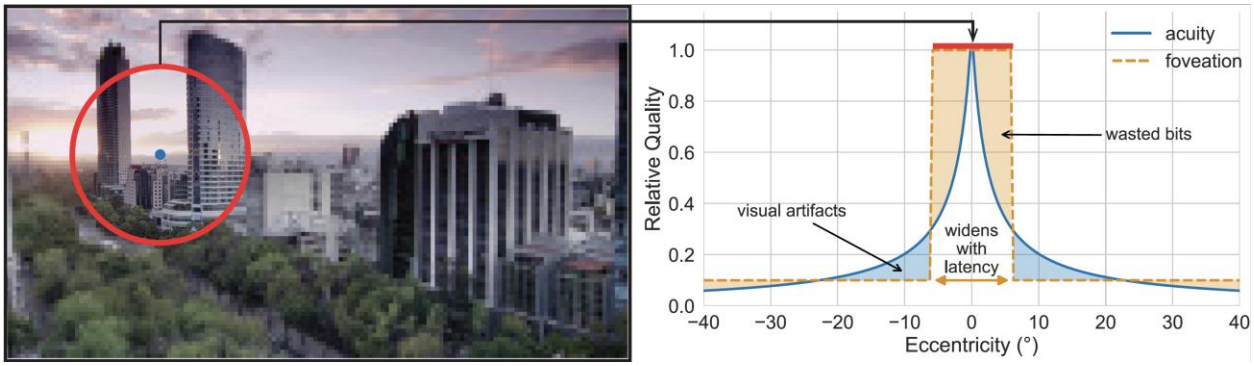

Image credit: Luke Hsiao, Stanford University

The team's "gaze-contingent" foveated compression and streaming system has a total latency—from eye movement to generated image—of about 14 milliseconds. Current VR systems have an end-to-end latency between 45 and 81 milliseconds. And this doesn't even take into consideration the time it takes to transmit data to the network. The team's approach includes time for eye-tracking (~1.5 ms), encoding and decoding video (~5 ms), HDMI video output (~3.5 ms), and physical transition of the LCD (~4 ms), but similarly not time to transmit data via a network.

Approaching reality

The scientific principle behind the technology is based in biology. Human vision encompasses a massive field roughly 220° wide by 135° tall. And yet, only a small region, just 1.5 percent of

that area, requires the finest detail—the fovea. Outside that small region, the rest of the visual field grows steadily less precise the farther from the fovea it gets. With foveated compression, however, relatively slower computer systems mean the area of foveated vision must grow wider to retain critical detail within the viewer's gaze as the eyes move across an image.

The authors have applied this principle to conserve precious bandwidth. That is, the area of greatest visual acuity (which therefore requires the highest resolution) is confined to a very small area of the VR screen where the user's eyes are focused.

"The tighter we keep that area of high resolution, the less data we use," explains Luke Hsiao who, as a doctoral candidate in the lab of Keith Winstein, was first author of the paper.

The size of the area of focus is closely tied to the computer's ability to track the wearer's eye movement. It must be wide enough to accommodate the delay in the computer's computational ability to keep up with the viewer's darting eyes. This delay is known as latency. The shorter the latency period, the smaller that area of critical focus can be.

"The bottom line is that *if* you can get the latencies down to the 15-millisecond level, you can achieve a roughly five-times reduction in bitrate over the current state-of-the-art systems using foveated compression that is imperceptible to the user," Hsiao said. "That's the goal of immersive technologies."

"This is equivalent to roughly 20 years' worth of progress in compression technology," Winstein, the senior author, said. "It would get us closer to making 'retina-quality' VR video streaming over the internet practical."

Promising prototype

The prototype employs a two-stream approach in which two versions of the current frame get compressed. The first frame, the a less-precise area outside the fovea, is compressed at significantly lower resolution. The second frame remains at full resolution but gets cropped to a small area nearest the gaze location.

The server then encodes both frames and sends them back to the client where the process is reversed. The client upscales the lower-resolution background frame to display size. It then decodes the second high-resolution foveated frame precisely within the viewer's gaze. The two images are overlaid and blended seamlessly to produce a single foveated frame.

The approach is more involved computationally than current VR technologies, but the bitrate is reduced considerably. User-study results imply there is "cliff" between ~80 ms and 45 ms latency range of today's best systems that led to user annoyance as the foveated region snapped into place on screen—the systems couldn't keep up with the user's eye leading to a less-than-immersive experience.

In their studies the researchers found that only at the lowest achievable latency were a substantial savings in bandwidth realized. Latency emerges as an important factor in user experience, the researchers say, and the research community should further investigate the tradeoffs between lowering latency in gaze-contingent video transmission and resulting bitrate reductions. Algorithmic approaches, they say, might not be necessary, if VR systems can be optimized.

“From these studies, it seems that 14 milliseconds is good enough in terms of system latency, but that gives the network *zero* time to send data,” Hsiao adds. “Next, we'd like to better understand the maximum latency budget we have to perform compression while still providing that truly immersive experience users expect.”

The research team also included graduate student Brooke Krajancich and professors Philip Levis and Gordon Wetzstein.

Luke Hsiao, Ph.D. Image credit: Nicholas Chiang

*The **eWEAR-TCCI awards for science writing** is a project commissioned by the **Wearable Electronics Initiative** (eWEAR) at Stanford University and made possible by funding through eWEAR industrial affiliates program member Shanda Group and the **Tianqiao and Chrissy Chen Institute** (TCCI®).*