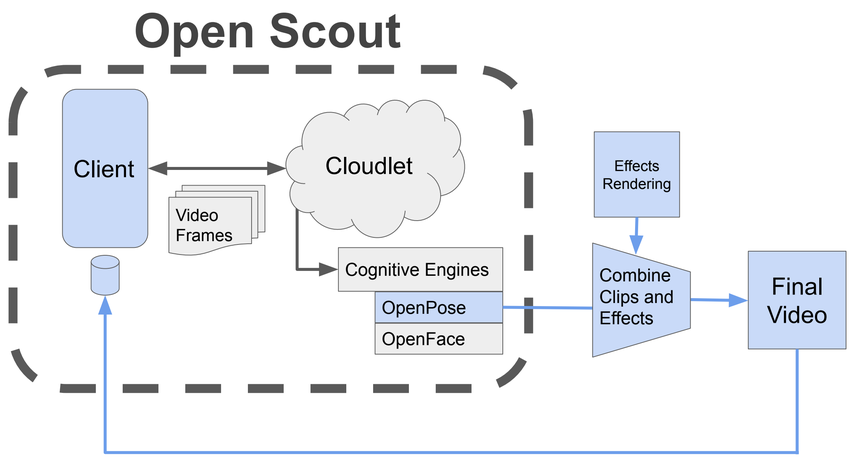
Media Composition on the Edge

George Ralph and Jim Blakley (Mentor)

Background

- Many social media platforms allow editing/composition of media on-device
- Popular filters often involve neural networks, facial tracking or augmented reality
- These effects can be compute intensive
- This project moves that heavy compute to a cloudlet

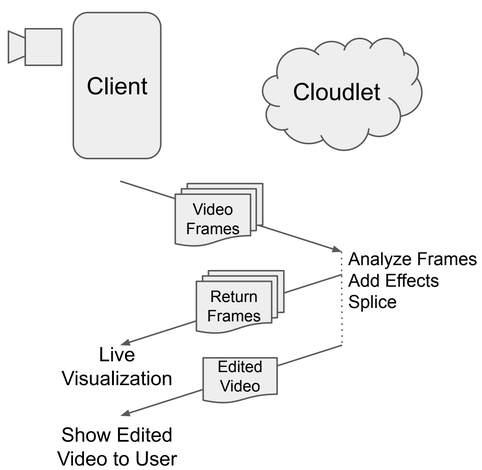


A block diagram of the system. Components within the dotted rectangle make use of the OpenScout framework. Blocks colored blue represent components which have been added or significantly modified.

Architecture

- Video analysis and editing is offloaded to a nearby cloudlet
  - Enables more advanced analysis/graphics
  - Pose tracking with OpenPose
- OpenScout is used to manage flow control between client and cloudlet
  - Built on Gabriel
- Live frames with effects applied are returned to the client in real-time
- When capture is completed a video is rendered and downloaded to the client via HTTP
  - Allows non-linear splicing of clips

Input video frame (left) and output video frame with pose identified and 3D effects composited (right)



Timeline of communications between the client (left) and the cloudlet (right). Return frames are sent in real-time, while the edited video requires more time to be constructed.

Results

- Final implementation operates at frame-rates near 20 FPS
- Effects applied to frames and returned to device in real-time
- Enables mobile devices to process videos with pose models too large for mobile hardware (2.2 GB)[1]

Lessons / Challenges

- Pose can be relatively easily interpolated between frames
- Server configuration is harder than it appears
- Android Development is easier than it appears
- Python multithreading is limited and has unfortunate blocking semantics
- Cloudlet proximity is critical to latency

[1] https://cmu-perceptual-computing-lab.github.io/openpose/web/html/doc/md_doc_05_faq.html