M&T Integration Lab 2021/22
Ultra-Efficient Video Chat - CIS401 Spring Report
Raymond Mason, Shaya Zarkesh, Karthik Macherla, Ashwin Nathan

Value Proposition

Need: Video chat has become an integral part of how we work and socialize, accelerated by the recent COVID-19 pandemic. We use video chat all the time and everyone knows how frustrating a bad video call with stuttered frames can be. Video calls are also notorious for quickly using up all your data. Other than just the cost, it will take significant time for fast internet to be available around the world, especially in more remote areas. Beyond the consumer experience, video chat in the form of conferences via Zoom or Microsoft Teams takes up significant company bandwidth. Ultra-Efficient Video Chat (UEVC) does exactly what the name suggests and provides an entirely new creative approach to video calls that uses 10x less data.

Solution: UEVC is a video compression algorithm that leverages the fact that most video calls are of human faces by sending facial expression data between users rather than the industry standard of compressed general-form video. Our MVP highlights how smooth and accurate UEVC can be while utilizing 10x less data.

Value Proposition: Make good video chat possible on low bandwidth so users can save data and video chat wherever and whenever. Allow for smooth video chat in fast-moving situations such as when someone is walking around buildings and elevators. Allows people in situations with slow internet to access video chat. Furthermore, for enterprise video chat software developers help improve call quality and reduce network congestion.

Stakeholders

Retail End Users: Want to video chat family and friends across the world using desktop or mobile applications. Interested in maximal quality for minimal data usage (on mobile). Incredibly low switching costs, high elasticity to price and high network effects.

Incumbent Video Chat Developers: Existing software companies whose primary product or one of the primary features is a variation of video chat. For example, Zoom, Skype, Whatsapp, Facebook Messenger, etc.

Commercial End Users: Want to use video chat for live conferencing for business purposes primarily using desktop applications. Interested in maximal quality and reliability, willing to use more data or pay to facilitate. High switching costs, low elasticity to price.

Market Research

Our research survey indicates strong demand for video chat, with most people surveyed using video chat highly regularly with more than 80% using it once every three days. Furthermore, people spend

significant time on video chat with average use times concentrated around 30 minutes. Quality is an incredibly important factor for users with most ranking its importance 9 out of 10. These data points indicate that there is strong demand for video chat with frequent usage and lengthy calls making an efficiency improvement worthwhile. However, most people (at least at Penn in the US) reported having an unlimited data plan. Thus it was no surprise that data usage was not ranked as an important consideration for most video chat users. While this may reduce the weight of the initial value proposition of saving data, we believe that the focus on quality by users can be capitalized on by providing a more efficient video chat that can offer high quality when only low bandwidth is available. Furthermore, data providers like AT&T may be interested in reduced data usage helping to lower network congestion and could urge enterprises that offer video chat applications to switch over to a more efficient protocol.

Target Customer Segment

We think there are two main routes we could take for our app. First, a direct-to-consumer model where we have an app on the app store that people download and use to video chat with each other. We are skeptical of this model, however, since it has a major cold-start problem: there will have to be a critical mass of people that actually have the app so that people can video chat using it.

The more promising revenue model is a technology licensing model to existing video chat companies, like Apple or Zoom. We can promise them better user experience for their users (less choppiness because of less required bandwidth), and reduced data usage for their end users. Eventually, we hope to be bought out by one of these companies.

Market Size Estimation

Top-Down: According to Fortune Business Insights, the Global Video Conferencing market size was estimated at $6.3Bn in 2021. Given the concentration of this market in several large players like Microsoft Teams, Zoom, Skype, etc. it is realistic that through only a few licensing agreements a large segment of the market could be captured (the above three alone would supply more than 50% of the above $6.3Bn TAM). As such, assuming the ability to charge both a fixed and royalty based variable licensing agreement summing to approx. 0.2% of revenue would facilitate a $6mm+ revenue opportunity with just these three providers.

Bottom-Up: Assuming based on survey results (shown in appendix), conservatively based on the sample group being UPenn students, that an average global person video calls once every three days for on average 30 minutes. Thus, across a year on average a person video calls for 60 hours (360 / 3 * 0.5). Assuming that our licensing agreement could allow us to reach users in US and EU (Whatsapp, Zoom, Teams, etc.) would imply 330 + 450 = 780mm people of which conservatively 60% would use a video chat service regularly as outlined above implying 470mm potential users. In terms of economic value, reduced data spending would save each of these users at least $0.10, although likely more though customers may not be willing to pay explicitly for these savings. Video chat providers may be willing to pay at least such an amount per user in a variable licensing agreement, resulting in a $4.7mm+ revenue opportunity.

Growth: Accelerated by the recent COVID-19 pandemic, the use of video calls especially in the workplace has become even further widespread. As we begin to emerge from the pandemic most office workplaces seem to be opting for a hybrid model maintaining a strong reliance on smooth and

clear video calls. According to Fortune Business Insights, the video conferencing market is expected to grow at a 11.5% CAGR reaching $14.6Bn in size in 2029. Given our goal to partner with leading technology firms that have strong network effects and resources to support growth, we believe our revenue streams from licensing would be reliable and could grow at the aforementioned rate.

Competition

Much work has gone into reducing bandwidth of streamed video. However, to our knowledge, no significant effort has been made to use a model of a face for lower video-cost video chat in a production application. NVIDIA research has released a few papers using AI-based methods to reduce the data usage of a video of a face, but they have not commercialized the idea, and using AI end-to-end for this task is likely to not be practical in a realtime setting on commercial devices.

Interestingly, Apple has almost tackled the same problem, but as a novelty feature. They allow a video chat where a person's face is replaced with their Animoji. This method is not to reduce data usage, as it still sends video frames, although Apple could easily apply this technology to reduce bandwidth by only sending features of the Animoji model without the video frames.

Intellectual Property

We are building a unique technology using image processing methods combined with video heuristics to build an efficient real-time engine to model and warp a face. The model is based on a series of calibration images. On each frame of the video, nearly 500 key points of the face are detected and sent over the network. The receiver in turn combines the calibration images and utilizes Delaunay triangulation to warp the calibration images into a face that mimics the current expression. The overall algorithm, in particular the image reconstruction phase, constitutes IP. We are attempting to file a process patent for the image calibration and reconstruction algorithm.

Revenue model

For a licensing model, we would benchmark our technology against existing incumbent strategies and demonstrate both the improved quality performance (reduced choppiness) as well as lower data usage. We would reach out to companies like Apple, Zoom, and Google directly and offer to help with integration into their existing video chat services. We would have a flat rate for developer support with integration, as well as a per-gigabyte data cost (something extremely small like 1 cent per gigabyte).

Cost

Costs are exclusively in the upkeep and improvement in the image reconstruction algorithm – since we are licensing the technology, we do not bear data costs, and since the algorithm is proprietary, it does not make use of any paid services. In order to turn this into a fully-fledged product, we'd like to hire at least two developers with expertise in computer vision.

Appendix (Survey Data)