Computer Vision, Extended Reality, and the Next Evolution of Mobile Devices

Jack Landers, March 20, 2024

A Future Through the Lens

Extended reality is the mobile technology of the future as it enables the display of your devices to be seamlessly integrated with your everyday vision.

In the 2000s, Apple released revolutionary mobile devices, which have led to over a decade of the iphone dominating the technology market as a necessary personal item. Last month (February 2024,) they added a new product line to their stores, a headset. This, along with massive investments from other titans of technology, mainly Meta, gives us a glimpse into the future of mobile communication and media devices.

Computer Vision, the technique that lies at the heart of these new technologies, is the use of programming to interpret camera data such that a computer can, like us, see. It has already begun to revolutionize other industries with its limitless applications such as robotics and automation.

A key example is Tesla, who have risen to compete with century old car companies, in only a few years, by harnessing image processing in both its products and manufacturing processes.

As the hardware develops, extended reality products will be able to effectively harness computer vision for mobile devices and make us further interconnected through online applications. Phones and social media have led to massive strides in the frequency at which we communicate. To previous generations it would be surreal that we can share information with one another without time or space as limits, and it will only continue to get faster. I aim to consider how our behavior as a society will change as our interactions with the virtual world do, but this begins with diving into how these devices function; only then can we start to realize how they might affect us.

Could a Computer Really See?

With modern solutions the concept of a computer with eyes is no longer only a figment of science fiction.

An algorithm uses machine learning where programs are trained by learning from experience, in order to analyze data and spot trends in images. It's an algorithm that is rewarded for completing tasks and adjusting itself to achieve a higher accuracy, developing a neural network which remembers successful techniques. It was originally created to train artificial intelligence models, but has since been repurposed for further applications.

Computer vision techniques are trained for:

- **Detection**, which identifies where in the image an object is located. This is done by training a neural network with the grayscale or black and white versions of an image so that it can detect patterns in defining contours and similar shapes.

- **Segmentation**, that separates the parts of an image that are each identified. It does this with a neural network that looks for patterns in the colors and shading, measuring hue and brightness.

- **Classification**, which attributes labels to different objects in the image with meta-learning where a program is pre-trained to identify different objects using a dataset of images.

- **Facial Recognition**, used to determine whether there is a face, from an album, present within an image. It does so using kernels which compare patterns in neighboring pixels.

- **Particle Swarm Optimization**, which uses images taken at similar time periods and compares them so that the identified objects can be used as reference points to reconstruct blurred areas from video that were not previously usable.

When a computer vision program classifies objects it is determining a weight for its confidence. This is a score from 0 to 1 that represents how alike an object is to images of it in the dataset that the neural network has learned from, because it can never be definitively certain. Instead if the weight is over a certain threshold, say 0.9, then we accept that it is sufficiently confident that the object has been detected.

Figure 1: An example of a Computer Vision software called YOLO, using detection to box objects, and classifications to label them, mapping the image so that it can be processed.

Where is This Applied in Practice?

Tesla

Tesla, and other automated vehicles, function using a combination of computer vision, and light detection and radar (LIDAR). The system within these cars is designed to detect and 3D map all objects around the car through depth perception. This information is interpreted by neural networks rewarded for safe driving that matches datasets with years of driving videos from consumers' cars. With these neural networks, there is enough output data about the surrounding environment to assist drivers by taking control of the wheel and driving for them.

On top of this, Tesla uses robotics to expedite their manufacturing process. Cameras relay data from the entire factory floor that is interpreted with computer vision and artificial intelligence processes that can be used for every aspect of the production process from quality control to industrial engineering.

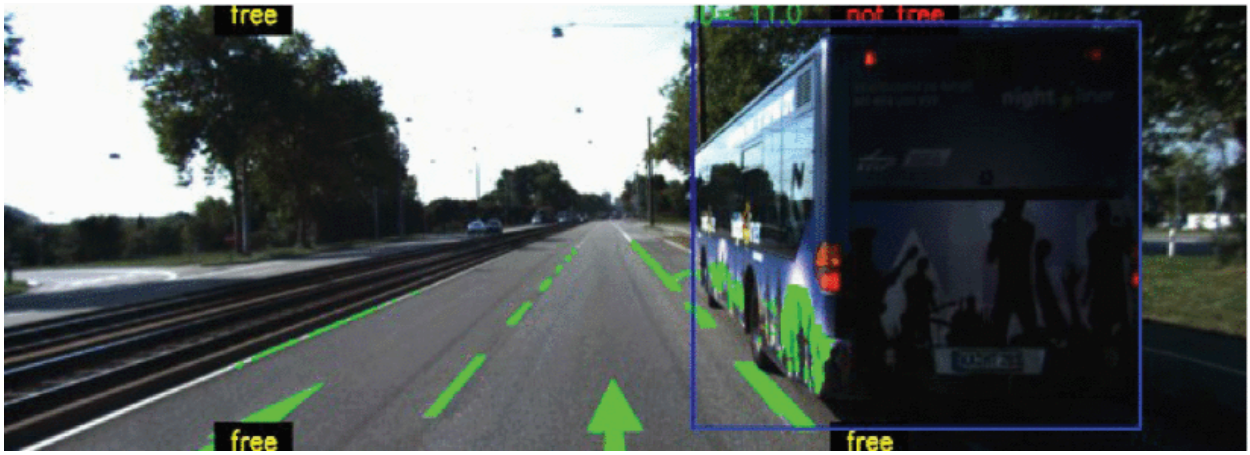

Figure 2: The image from computer vision and LIDAR data as an automated vehicle drives on a highway, with labels indicating where the system believes it to be safe to drive, augmented on top

Such effective uses of computer vision have massively progressed both the autonomous vehicles and the robotics industries alike. In part, their success has been a reassurance for mobile device companies to invest in researching augmented and virtual reality applications for their market.

Meta

Meta have already established dominance in the industry of extended reality with their significantly cheaper Quest headsets. This was a long term plan executed by Facebook when they acquired Oculus (the original Quest developers,) in 2014, and then rebranded to become a virtual reality company in October 2021. This was an extreme change for the company that had established itself as the most popular social media site in the world. Although, it is only a VR headset, transporting users to captivating virtual landscapes, at the cost of an awareness of their surroundings. The goal is to virtualize spaces that users can view and interact with, to make limitless possibilities accessible from a user's bedroom.

Apple

Now, they are competing with Apple, which recently released a mixed reality headset that seamlessly augment the real world with digital overlays, allowing users to see and interact with virtual objects in their physical environment. High-resolution displays coupled with a camera, create an immersive experience similar to a phone, but are designed for capturing spaces or using apps in three dimensions unlike any devices we've seen. While there are drawbacks, such as comfort and battery life, the Apple Vision Pro focuses on user experience and positions itself as a strong contender in the burgeoning world of spatial computing, despite costing significantly more than competition.

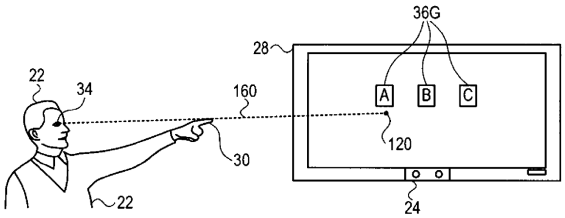

FIG. 11

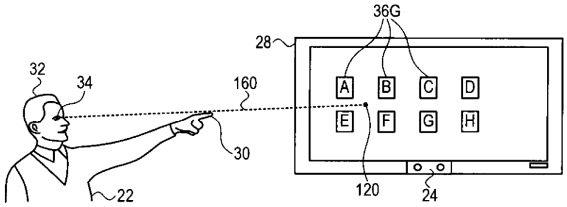

FIG. 12A

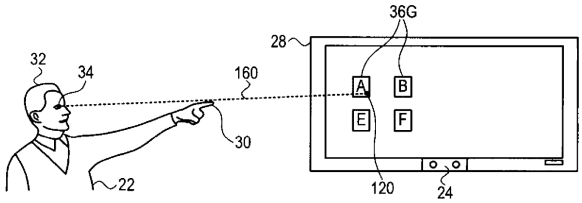

FIG. 12B

Uniquely, Apple applies the use of hand gesture recognition, a technique that demands a superior processor. Patents suggest they intend to combine hand signals with eye positioning to expand the range of possible inputs even more in the future, such that users can point to the virtual windows. In turn it rewards the user with an especially new technique for interacting with the device, that can increase the frequency of inputs from those we are currently used to, like a keyboard or touchscreen.

Figures 11-12: A schematic from "Gesture Mapping in a 3D Environment," a patent filed by Apple Inc.

At the moment, virtual reality headsets are most popular among the niche gaming market, however, Apple's entry into the market as an augmented "mixed" reality headset foreshadows a shift in the consumer base and the applications used towards communication and social media.

How Will This Technology Evolve?

Computer vision technologies are developing at rapid rates as it is applied to a plethora of industries. Established tech companies will continue to research and market extended reality. As well as this, companies will adapt to produce headsets at a greater scale, reducing prices. With these considerations, I believe this will lead to headset computers becoming normalized, day-to-day consumer products that we all require, replacing our phones.

An evident barrier of entry for consumers is the stigma surrounding interactions in a virtual space, but its practicality has begun to outweigh the social negatives and turn consumers' opinions. Accessing a mobile device that is integrated with your vision, further improves communication, because it tempts users to stay permanently connected to their device as they perform tasks. When augmented information can be consumed simultaneously with the natural world we experience the barrier between virtual and real beginning to fade, just as it does when we start endlessly scrolling on our phones.

A current example of this is Meta's development of virtual reality facetime, where users' virtual avatars are designed to be photorealistic. This has yet to reach consumers as it requires complex face scans, but Zuckerberg aims to introduce it to the market within the near future.

What has to Happen First?

On the other hand, a major setback in the development of augmented reality headsets, is that computer vision requires an overwhelming amount of processing, leading to them producing too much heat on users' heads and making for an uncomfortable experience. In response, the National Research Foundation of Korea recently found multicore fibers are a solution, which would transport the output imagery from a different part of the headset, rather than the face. This would relieve the user's face of heat from local processors and projectors located in the front of the headset.

It proved effective at reducing the temperature, but this comes at a cost to resolution. With further research and development, it could be solved in the near future, but for now, direct projection, as is used in the current industry standard headsets, is superior.

At NVIDIA, the most significant developments in hardware are occurring where new large language models have created a massive spike in demand for their graphics processing units. This has enabled them to reinvest in researching new processing techniques that will accelerate the speed and efficiency at which neural networks can learn, another solution to these problems. The demand is so high that it has also pressured AMD to develop their own GPU chip in order to compete, that could be specialized for devices like headsets.

NVIDIA is also researching new computer vision techniques, pushing the limits of the software to improve. This is particularly in areas of emotional interpretation. With perfected neural networks, facial detection can be applied to interpret other people's emotions even more effectively than we might. In addition to emotions, researchers have found that people's posture and gaze can be analyzed such that the algorithm can detect their state of drowsiness, cognitive load, and distraction. Such techniques could be paired with large language models produced by

Google for Alexa and OpenAI for GPT, and used for a variety of applications in automating aspects of social interaction.

Predicting the Effects

As this technology becomes increasingly realistic, it is important to ask ourselves, is this the future that we want?

Further empowering our mobile devices could be compared to opening Pandora's box. The use of social media and video games is an addiction that has already become so normalized where spending countless hours a day online is not abnormal.

Connecting to virtual worlds or metaverses may bring us closer together and entertain us further, but this comes at a cost. From within a headset we stray further away from the natural world and this could potentially have consequences that we must consider.

The development of these technologies further immerses consumers into the already massively addictive internet that is reshaping our mental health. In the National Library of Medicine, Dr. Mehendale considers how this leads to a development of mental disorders and weighs this against the support that rehabilitation applications of virtual reality can supply. Overall, they come to the conclusion that there is potential for mobile computer vision devices to have either a profoundly negative or significantly positive effect on society's mental health depending on the users and the applications of the technology that they choose to use.

In government, regulators continue to push back against social media companies, where influences from foreign surveillance states pose a threat to misinformation and cybersecurity, threatening democratic values. To oppose this, technology companies are in no short supply of funding that can continue to bankroll lawyers that lobby and fight lawsuits against any counter efforts, in what appears to be a never ending cycle of litigation.

In Conclusion

Following trends in industry, we can see how the introduction of computer vision techniques to mobile devices will bring limitless applications that could both benefit and harm consumers depending on their use cases. Successful companies in social media, artificial intelligence, and robotics are all incentivised to research and invest in its future. This has started to lead to heaps of progress in object detection, facial recognition, and depth perception. AR headsets are the next

generation of mobile devices that will harness such techniques and integrate the virtual with the real world. This expansion will evolve how we communicate with one another, consume media, and manage information. There are still technical limits to overcome, but industry recognizes the technology's potential and the largest companies in the world are invested in computer vision, making its success inevitable.

Sources

1. **Y. Zhang and X. Zheng, "Development of Image Processing Based on Deep Learning Algorithm,"** *2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, Dalian, China, 2022
2. **S. V. Mahadevkar *et al*., "A Review on Machine Learning Styles in Computer Vision—Techniques and Future Directions,"** in *IEEE Access*, vol. 10, pp. 107293-107329, 2022
3. **S. Veni, R. Anand, D. Mohan and P. Sreevidya, "Lidar Based Lane-Change Assistance For Motor Vehicle,"** 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021
4. **C. Riordan and C. O'Donnell, "Fusion of LiDAR and Computer Vision for Autonomous Navigation in Gazebo,"** SoutheastCon 2023, Orlando, FL, USA, 2023
5. **P. Rakhmetova, A. Bektilevov, B. Kopenov, M. Zhamuratova, S. Yussupova and T. Toktamyssova, "Study of Computer Vision for Manipulators,"** 2023 4th International Conference on Communications, Information, Electronic and Energy Systems (CIEES), Plovdiv, Bulgaria, 2023
6. **C. A. L. Waechter, D. Pustka and G. J. Klinker, "Vision based people tracking for ubiquitous Augmented Reality applications,"** 2009 8th IEEE International Symposium on Mixed and Augmented Reality, Orlando, FL, USA, 2009''
7. **J. -S. Lee, S. -H. Cho, W. J. Choi and Y. -W. Choi, "Multicore Fiber Relay Optics for Comfortable See-Through Vision in Augmented Reality Display,"** in IEEE Photonics Technology Letters, vol. 36, no. 4, pp. 219-222, Feb.15, 2024
8. **Usmani SS, Sharath M, Mehendale M. Future of mental health in the metaverse.** Gen Psychiatr. 2022 Jul 22
9. **Nuri Murat Arar, Niranjan Avadhanam, Nishant Puri, Shagan Sah, Rajath Shetty, Sujay Yadawadkar, Pavlo Molchanov, "Neural network based facial analysis using facial landmarks and associated confidence values,"** Google Patents, US, Aug. 27, 2020, US20210182625A1

10. **Feng Hu, Niranjan Avadhanam, Yuzhuo Ren, Sujay Yadawadkar, Sakthivel Sivaraman, Hairong Jiang, Siyue Wu, "Gaze detection using one or more neural networks,"** Google Patents, US, Aug. 19, 2019, US20210056306A1
11. **Eyal Bychkov, Oren Brezner, Micha Galor, Ofir Or, Jonathan Pokrass, Amir Hoffnung, Tamir Berliner, "Gesture detection in a 3d mapping environment,"** Google Patents, EU, Aug. 21, 2019, EP3527121B1
12. **Mark Zuckerberg: First Interview in the Metaverse | Lex Fridman Podcast #398** https://www.youtube.com/watch?v=MVYrJJNdrEg
13. **Matthew Johnston, "10 Biggest Car Companies,"** Invesopedia, Dec. 19 2023 https://www.investopedia.com/articles/company-insights/091516/most-profitable-auto-companies-2016-tm-gm.asp#toc-8-tesla-motors-tsla
14. **Isaacson, Walter. "Elon Musk."** Simon & Schuster, 2022
15. **Sheera Frenkel, Mike Isaac and Ryan Mac "How Facebook is Morphing into Meta"** New York Times, Jan. 31, 2022 https://www.nytimes.com/2022/01/31/technology/facebook-meta-change.html
16. **Josh Constine, "Facebook's $2 Billion Acquisition Of Oculus Closes, Now Official"** TechCrunch, July 21, 2014 https://techcrunch.com/2014/07/21/facebooks-acquisition-of-oculus-closes-now-official/
17. **Nat Rubio-Licht, "Google is Adding Empathy to it's Smart Speakers"** The Daily Upside, Sept. 25, 2023 https://www.thedailyupside.com/technology/big-tech/google-is-adding-empathy-to-its-smart-speakers/