Object and Facial Recognition in Augmented and Virtual Reality: Investigation into Software, Hardware and Potential Uses

Erin Schulte
Kennedy Space Center
B.S. Visualization
AVR Intern: 2017 Fall Session
11/14/2017

Object and Facial Recognition in Augmented and Virtual Reality: Investigation into Software, Hardware and Potential Uses

Erin E. Schulte
Texas A&M University, College Station, TX, 77843

Acronyms

2D – Two Dimensional
3D – Three Dimensional
AR – Augmented Reality
HCI – Human-Computer Interaction
MCI – Mass Casualty Incident
RAM – Random-Access Memory
SDK – Software Development Kit
VR – Virtual Reality

I. Introduction

Augmented and virtual reality are two technological concepts whose origins can be traced back to the beginning of the 19th century when painters attempted to create panoramic artwork that could provide a life-like, immersive experience for their viewers. An investigation by Charles Wheatstone in 1838 revealed how the human brain processes two-dimensional (2D) images from each eye into a single three-dimensional (3D) image, which means that if someone views two stereoscopic images side-by-side, the viewer will experience a sense of 3D depth that a regular 2D image cannot provide[4]. The discovery of how the human brain perceives and processes depth coupled with a desire to create immersive visual content has laid the groundwork for modern Augmented Reality (AR) and Virtual Reality (VR) technology development.

As AR and VR grows in popularity, and more researchers focus on its development, other fields of technology have grown in the hopes of integrating with the up-and-coming hardware currently on the market. Namely, there has been a focus on how to make an intuitive, hands-free human-computer interaction (HCI) utilizing AR and VR that allows users to control their technology with little to no physical interaction with hardware. Current attempts at such a user interface include speech recognition and motion tracking technology, which are commonly used in mobile phone applications and video games.

Computer vision, which is utilized in devices such as the Microsoft Kinect, webcams and other similar hardware has shown potential in assisting with the development of a HCI system that requires next to no human interaction with computing hardware and software. Object and facial recognition are two subsets of computer vision, both of which can be applied to HCI systems in the fields of medicine, security, industrial development and other similar areas.

Already, companies have found ways to utilize facial recognition, object recognition, VR and AR in the workplace to increase the efficiency of their businesses. This technology has shown stunning rates of development and has proven itself in many situations to be an efficient and convenient way to seamlessly integrate technology into the everyday lives of individuals around the world. In time, object recognition, facial recognition, AR and VR will develop to the point where they can be utilized in most, if not all, work environments.

II. Initial Investigation

When investigating quickly growing areas of technology, it becomes a requirement to periodically check scientific publications in order to see what other groups are currently researching, prototyping, and how they are approaching their studies. With my first focus being solely on object and facial recognition software, I planned out a method of checking new releases of scientific papers every week so that I would remain up to date with new discoveries, methods of development and overall applications of object and facial recognition.

As I moved through my investigation, it quickly became apparent that many different organizations are trying to find a simplistic way to create and seamlessly integrate object and facial recognition into AR and VR environments, though everyone's approach differed slightly in execution depending on their resources and the field in which they wished to deploy their technology. The versatility of object and facial recognition can be seen in the wide variety of

fields in which this technology can be applied. The following are a few of the many applications that have been created with object and facial recognition: 3D object recognition for robotic grasping systems in a factory environment[1], a wearable face recognition system for visually impaired users[8], and 3D vision-based detection system for broccoli heads in a field to assist with efficient harvesting practices[5].

Despite the varied applications and development approaches for object and facial recognition, a few core mathematical concepts can be seen in use throughout the realm of this technology. The Viola-Jones algorithm, which was proposed in 2001, "brings together new algorithms and insights to construct a framework for robust and extremely rapid visual detection."[10] This algorithm has formed the basis of facial recognition software and can be seen in use across a number of different research projects. One of the core functions of the Viola-Jones algorithm is the detection of Haar features, which are regions of almost every human face that share similar patterns of light and shadow. For example, the bridge of a human nose is brighter than the upper-cheeks. The upper-cheeks are brighter than the eye sockets, and so on. There have been many variations on the Viola-Jones algorithm, but the core mathematics have remained the same since it was first proposed.

Other common mathematical concepts that are frequently spotted when investigating facial recognition are Eigenfaces[11], Fisherfaces[7], and Laplacianfaces[3]. These three methods of processing images aim to identify if images contain a face and in what area of the image the face is located. But the process with which they identify faces is quite different, even if the concepts that these methods are built off of are fundamentally the same. As my investigation continued, it became clear that there are many ways to create programs that detect and recognize faces, as well as objects, and therefore attempting to look in-depth into each possible path to success was not an efficient use of time. Looking into comparisons between each method of object and facial recognition to find a frontrunner was an unsuccessful venture as well, since all methods have benefits and shortcomings that make each different path a viable possibility for the best overall method depending on how one intends to use the technology.

After the realization that solely looking at how to create object and facial recognition software packages was not effective in determining which method produced the best results, I determined that I should switch my focus from searching primarily for information on the development of this technology and instead begin to assess current software packages for object and facial recognition that have the potential to be compatible with AR and VR environments. As the primary goal of my investigation was to locate the best object and facial recognition technology for use in AR and VR, switching from investigating the complex methods of creating software to investigating existing software was the most viable solution as to how I should budget my time to garner the best results.

III. Software Solutions

A variety of object and facial recognition software packages currently exist and have been released to the public, but most of the current applications of said software are focused more on entertainment and low-level security clearance than issues which hold more weight, such as patient identification in medical settings, high-level security clearance, emergency response team assistance, and other tasks that hold the possibility of impacting human lives if the technology does or does not function correctly. For example, the iPhone X utilizes, to some

extent, facial recognition technology designed to make their phones more secure by using an individual's face as a tool for unlocking their phone. Popular phone applications, such as Snapchat, use facial tracking to identify where someone's face is in a camera view and activates an AR overlay to apply a filter over the person's face. Many websites offer services that can help tag uploaded images by detecting objects, colors, and individuals in said images and then categorize them based on a certain set of rules established by the user. These types of applications prove that object and facial recognition are viable technologies that people have expressed interest in using, but only a few developers have focused on advancing the technology for use in situations where the software's accuracy impacts more than just a social media photograph or locking and unlocking a personal phone.

However, some object and facial recognition software packages exist that have shown evidence of developing beyond the commercialized uses that already exist for this technology. The following software packages were found to be the most promising with object and facial recognition and were assessed to determine how they would perform in situations which call for reliability of software, high accuracy in results returned and compatibility with numerous hardware and software platforms.

A. Luxand FaceSDK[6]

The software development kit (SDK) offered by Luxand demonstrated a high level of real-time recognition accuracy in the demo that they provided for those interested in viewing their software in action. Registering individuals into the facial recognition database was as simple as clicking on the face of a person in the camera view, typing in their name, and letting the computer do the rest. The more often the software saw a person's face, the faster and more effectively the artificial intelligence (AI) in the software package could recognize that individual again in another session. Once a user is registered, that user need not be registered again. The SDK did attempt to include an age/gender identification system, but it proved to be inaccurate on many occasions. The software is offered in a variety of programming languages and is also compatible with a wide array of operating systems and hardware.

B. VeriLook SDK[9]

Neurotechnology, the company behind VeriLook SDK, has created other recognition technology in the past, including eye iris and fingerprint identification. They have recently released their face recognition software, VeriLook SDK, as well as an object recognition package. Overall, the facial recognition software is accurate, but requires more interaction with the computer to register, identify, and read the individual to be identified from the image or video feed provided, which is a detriment when attempting to create innovative HCI platforms. The software displays the level of confidence in the individual identified, and also allows for the database to be updated as needed throughout the software's use. Their object recognition software proved to have somewhat low accuracy on the items we chose to test, but has received good reviews from companies and other users. Like Luxand, this software package is offered in a variety of programming languages and is compatible with most operating systems.

C. Vitruvius[2]

While not an object or facial recognition software platform, Vitruvius was assessed as a possible add-on for another software package in order to increase the accuracy of the facial recognition readings. While the other two software packages used 70 and 68 facial points, respectively, to track and identify the faces of its users, Vitruvius uses over 1,000 points to track the movement of faces. This software was developed primarily to be used with the Microsoft Kinect v2, but it can possibly be applied to other hardware platforms. Currently, the primary programming language for Vitruvius is C#, and it has proven compatibility with Unity3D, which is a commonly used program in the development of VR and AR applications.

These software packages all showed that they have the potential to be applicable in environments such as emergency medical situations and high-level access areas that require advanced security measures, but they all have not yet reached a level of accuracy that can be relied upon without expecting a certain margin of error. Based on the rapid growth of this field of technology, it is reasonable to predict that the software packages listed above will develop to the point where they will be usable in almost any situation in the span of a few short years.

Another factor that needs to be considered when assessing object and facial recognition software on a project by project basis is the size of the object or facial recognition database required for whichever task it is to complete. If one is using this technology to identify a single type of airplane part in a massive manufacturing factory, or one wants to allow a specific set of individuals to unlock a door based on whether or not their faces are in the database for that lock code, the database size for the facial recognition software is fixed and not very large. If one wants to use facial recognition with a large group of employees that is constantly changing or software is needed to identify objects in a previously unexplored environment, the software database will be much larger and constantly updating, meaning that the structure of the software itself has to change to accommodate a larger load of information flow. So, depending on the use of the object and facial recognition software, some existing program packages may work better than others.

Throughout the process of assessing the above software packages, evidence of other companies preparing to join the community of object and facial recognition developers could be observed on a weekly basis. New start-up companies would emerge consistently, but these companies have yet to release enough information to create an accurate assessment of the usability of their product. If a similar investigation as this one is made in the near future, a new company could emerge that can meet the need of having fast, reliable, and accurate computer vision recognition software.

IV. Hardware Requirements

After reviewing different SDKs that were manufactured for object and facial recognition, hardware was assessed in order to determine the best possible processor construction for this up and coming technology. As mentioned before, some object and facial recognition programs require large databases of facial and object templates for the program to refer to when identifying

someone or something. This requirement, in turn, limits the amount of database space to the amount of available RAM in a processor build. As the amount of RAM increases, the other hardware in the processor build will be impacted and may need to be switched out or replaced because the processor needs to be able to keep up with consistent RAM access. There are many possible hardware combinations that can work with object and facial recognition, though it is dependent on the user of the software to determine which hardware will best suit their needs since they have the best idea of how long and how frequently the facial recognition or object recognition software will be used, and how large the template database will need to be for their particular project.

In regard to integrating object and facial recognition into AR and VR platforms, it is up to the users of the recognition software to determine which AR and VR headset will best suit their needs. There are a variety of options, from the HTC Vive and the Oculus Rift in VR to the Microsoft HoloLens and Recon Jet Pro for AR. Each hardware system has advantages and disadvantages which will have to be weighed depending on how they will be used.

V. Potential Uses

As depicted above, object and facial recognition technology is versatile and constantly evolving, which allows this type of software to be a solution for a variety of problems in a number of situations. One of the most common uses that developers are targeting when creating this technology is using facial recognition to unlock doors, phones, and computers that contain sensitive information. While this technology already exists to some extent, it is not functioning at a level that is considered appropriate for the protection of classified information. Integrating facial recognition into systems that commonly use passcodes or keycards cuts down on the possibility of someone infiltrating a secured area, as it is easier to obtain a passcode or keycard than it is to trick a facial recognition program when it is innovatively created. Some companies, such as VeriLook, have advertised that their software can detect whether a person is "live" or "not live" when participating in real-time face recognition. This kind of capability deters the use of images of people's faces as a way to potentially trick facial recognition software. This potential use of facial recognition also assists those who have difficulty remembering passcodes or tend to forget their keycards, as the only thing they would need to unlock doors and their electronics is their face.

Object and facial recognition has also been considered a viable technology for use in search and rescue missions, particularly after natural disasters and during mass casualty incidents (MCI). During events such as earthquakes, wildfires and MCIs, paramedics and first responders on the scene need to perform primary triage on victims to determine how to best treat them. Due to the large amount of individuals who need to be treated after such massive events, and a lack of communication between first responders and the medical teams at hospitals located nearby, medical centers tend to become flooded with high levels of patient traffic and end up not being allocated enough resources to effectively treat everyone brought into their facility. Another problem that arises during these types of events is keeping track of individuals as they are moved from the site of the incident to surrounding care facilities, as well as identifying individuals on scene who do not survive.

During catastrophic events, such as the ones described in the above paragraph, the use of an AR headset equipped with facial recognition and other capabilities such as live video

streaming, a built-in camera, and similar features could drastically change how such events are handled. First responders on the scene could video chat with professionals at nearby medical facilities who may have a better idea of how to perform primary triage or treat a patient in desperate need, and can relay this information to the first responder as they work on the patient on-site. Resources can be better allocated to nearby hospitals because individuals not at the scene can receive information directly from the first response team and delegate resources as necessary. Similarly, individuals with less life-threatening injuries can be sent to hospitals farther away from the scene while people with immediate need for medical attention can be sent to closer hospitals, reducing the chaos that usually ensues in emergency rooms after such events. And, for the first responders who need to identify those who do not survive, facial recognition software can severely cut back on the amount of time spent around the deceased individuals. By identifying those who passed away quickly, it cuts back on the mental trauma inflicted on both the first responders and the individuals searching for loved ones, as they do not have to look at every body to see if it is their friend or family member.

An aerospace application for such technology includes the use of object and facial recognition in a VR environment during explosions and other similar incidents that happen in engine and rocket testing facilities. Before humans are sent to the scene of an incident, robots or drones equipped with object and facial recognition technology, as well as a 360 degree camera, can be sent in first. These machines can stream video footage in real-time to a VR headset at a safe location, so that users can interact with the site of the incident as if they were actually there while remaining at a safe distance. These drones and robots can then be remotely moved around the disaster site, using object recognition to identify potentially dangerous objects such as unexploded fuel tanks, shrapnel, and compromised structures. If there happened to be people on-site during the time of an incident, facial recognition software can identify anyone these robotic devices come across and send up a signal beacon on where the individual is located so that when first responders are allowed on the scene, they can find anyone who is injured quickly and efficiently. Or, in the worst-case scenario, the facial recognition software can identify who did not survive the incident.

Manufacturing facilities can equip their employees with AR headsets that contain object recognition software which has the ability to assist them in identifying parts when assembling complex components. These headsets can be equipped with live video feeds and instructional walkthroughs on how to properly assemble materials to further increase productivity. Boeing, an aerospace-based company, has already begun to integrate AR technology into its commercial aircraft manufacturing facilities and has reported an increase in the speed of its assembly line process. Boeing has also expressed an intent to continue to invest in AR and VR technologies in the future in order to investigate if AR and VR can be used to improve autonomous systems. The successful results already seen in Boeing's use of AR technology in its commercial airline manufacturing facilities suggests that if other aerospace companies like Blue Origin and SpaceX invested in similar technologies to the ones currently in use at Boeing, these companies could expect to see an increase in productivity in their spacecraft and launch vehicle processing. The processing of vehicles intended to travel through space is a lengthy, detail-oriented process that requires a highly-trained team of workers with impeccable communication skills, which is why AR would be a valuable asset to workers for not only communication, but quality-checking of work already completed.

VI. Conclusion

Investing in the creation of object and facial recognition software, as well as the hardware required in order to run said software, is a worthy investment for any group of researchers who want to have a hand in shaping the newest realm of HCI technology. Evidence of the viability and flexibility of object and facial recognition can be seen by observing the vast number of areas to which this technology has already been applied. The growth rate of this technology also demonstrates the interest that individuals around the world have in increasing its accuracy, reliability and applications far beyond what has already been accomplished. Already, this type of technology can be used on a small scale for entertainment purposes and low-level security, but when combined with AR and VR hardware, it has the capability to help save lives, improve the efficiency of manufacturing processes around the world, and put us on the path to an intelligent computer system which will change the course of human-computer interaction.

VII. References

[1]Chen, Chin-Sheng, et al. "Three-Dimensional Object Recognition and Registration for Robotic Grasping Systems Using a Modified Viewpoint Feature Histogram."*Sensors*, vol. 16, no. 11, 2016, p. 1969., doi:10.3390/s16111969.

[2]"Create Stunning Kinect Apps in Minutes." *Vitruvius*, vitruviuskinect.com/.

[3]He, Xiaofei. "Laplacianfaces." *Scholarpedia*, vol. 5, no. 8, 2010, p. 9324., doi:10.4249/scholarpedia.9324.

[4]"History Of Virtual Reality." *Virtual Reality Society*, www.vrs.org.uk/virtual-reality/history.html.

[5]Kusumam, Keerthy, et al. "Can You Pick a Broccoli? 3D-Vision Based Detection and Localisation of Broccoli Heads in the Field." *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, doi:10.1109/iros.2016.7759121.

[6]"Luxand - Face Recognition, Face Detection and Facial Feature Detection Technologies." *Luxand - Face Recognition, Face Detection and Facial Feature Detection Technologies*, www.luxand.com/.

[7]Martinez, Aleix. "Fisherfaces." *Scholarpedia*, vol. 6, no. 2, 2011, p. 4282., doi:10.4249/scholarpedia.4282

[8]Neto, Laurindo Britto, et al. "A Kinect-Based Wearable Face Recognition System to Aid Visually Impaired Users." *IEEE Transactions on Human-Machine Systems*, 2016, pp. 1–13., doi:10.1109/thms.2016.2604367.

[9]Neurotechnology. "VeriLook SDK." *VeriLook Face Identification Technology, Algorithm and SDK for PC, Smartphones and Web*, 9 Nov. 2017, www.neurotechnology.com/verilook.html.

[10]Viola, P., and M. Jones. "Robust Real-Time Face Detection." *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, doi:10.1109/iccv.2001.937709.

[11]Zhujie, and Y.l. Yu. "Face Recognition with Eigenfaces." *Proceedings of 1994 IEEE International Conference on Industrial Technology - ICIT '94*, doi:10.1109/icit.1994.467155