Emotion Detection

Calvin Knowlton and Jonathan Wang

Johns Hopkins University, Whiting School of Engineering

Baltimore, MD

ABSTRACT

For people with autism, reading facial expressions and emotions can be a very daunting task that takes a lot of time to learn and even more to master. Instead, we propose a way to harness the power of emotion detection technology to address such an important need on a very discrete level. Such a difficult problem however does not get any easier in the digital world, and limiting the size of our hardware makes this resource-heavy operation extraordinarily difficult. To overcome this, we have developed a lightweight, client-facing piece of software that connects to a suit of Google Cloud Platform tools to process freeze frames from a live video feed and identify facial expressions in real time. We envision implementing our software in smart glasses hardware, akin to Google Glass or Snapchat Spectacles. We envision our user wearing glasses that presents them with an augmented reality overlay of the emotion of the person with which they are interacting. We also will support a purely auditory experience, where the emotion will be presented to the user through bone conduction speakers embedded in the frame of the glasses. Our proof-of-concept implementation, however takes place on a smartphone – it periodically captures pictures using the rear camera and sends them to Google FireBase, which routes them to Google Cloud Vision API using FireBase Functions, parses the response, and relays the emotion back to the smartphone, where it is then displayed on the screen.

1. INTRODUCTION

Not everyone can easily discern facial expressions. For people with autism, reading facial expressions and emotions can be a very daunting task that takes a lot of time to learn and even more to master. This can lead to difficulty interacting with people, especially in face-to-face social situations and with people that they do not know well. As a result, people with autism can experience added stress and anxiety in their everyday lives. Current solutions for this problem include social skills therapy and do-it-yourself programs for parents. Social skills therapy requires a trained professional and is a very expensive and time-consuming process, as frequent meetings with the trained professional are required. Do-it-yourself programs for parents are cheap, since the parents simply need to engage in social skills therapy with their children, but such programs require intensive training on the parents' part, and are just as time-consuming as a more professional option, if not more. Additionally, there is a shortage of trained therapists. Children could be forced to wait up to 18 to begin receiving treatment months after receiving a diagnosis of autism. A simpler solution, smart devices and cloud computing technology can conveniently help people with autism read facial expressions and emotions. When used in concurrence with professional guidance and therapy, we believe that the total cost and length of therapy can be reduced, as well as increasing the effectiveness of said therapy. Our project aims to be a proof of concept of using cloud computing to streamline facial detection and incur less load on the smart device itself.

2. RELATED WORK

Currently there is ongoing research at the Stanford University using Google Glass to aid in facial recognition situations. Their implementation links the Google Glass to a smartphone device running an app that has pre-trained models to compute the emotion. The app recognizes eight different basic facial expressions and emotions. There are also three modes: "Free play", where the app simply communicates the emotion of anyone in the device's field of view, as well as two game modes, where the user can practice his or her skills recognizing facial expressions. After the pilot study, "the children's scores on a social-skills questionnaire indicated less-severe autism symptoms. Six participants had experienced changes in their scores large enough to move down one step in the severity of their autism classification, and 12 of the participating families reported that their kids made more eye contact." [1].

3. OUR SOLUTION

We intend our glasses to look and function more similarly to Snapchat Spectacles than to Google Glass since we envision our glasses seamlessly blending in with regular glasses, remaining discrete. Our implementation of the system also does not need the compute power and hologram feature that Google Glass has. That way, we can cut back on the device's bulk, and reduce the chance of strangers staring awkwardly

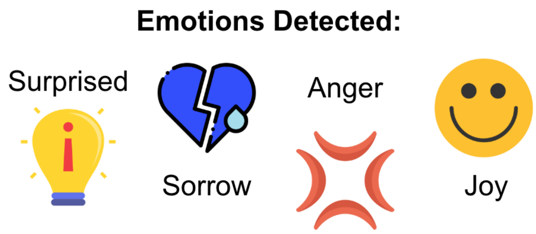

Figure 1: The different emotions detected by Google's Vision API.

at the user or other unwanted interactions that could interfere or complicate the therapy process. Most importantly, our solution makes use of a suit of Google Cloud Platform tools, including Google's Vision API which does all of the heavy lifting for facial detection and emotion recognition. Thus, the glasses will no longer have to be tethered to a smartphone for image processing. Additionally, the smartphone will not have to be constantly processing images from the glasses, so it will not use nearly as much battery power and internal resources. Removing the computing process from the hardware itself and moving it to the will also reduce the cost of the overall system. Instead of having to spend $1,500 USD on Google Glass [2] and having to purchase a smartphone to pair with it, users can spend about $150 USD on our hardware, which will be similar to Snapchat Spectacles [3], and nothing else.

4. DESIGN

As a stand-in for our glasses and in order to prototype, we built a simple mobile app using the Ionic framework that uses the camera to take an image every few seconds. The app then sends a compressed version of the image to a FireBase bucket. When it detects a new image, a FireBase Function executes, sending the picture to the Google Cloud Vision API for analysis. The Vision API will identify faces in the image and return their location and percieved emotion. FireBase receives the analysis and stores the data in a Firestore NoSQL database. Listening for this update, our app receives the information from the analyzed image and relays the results to the user in real time with a sound and an icon. For the icons, we chose a set of easily distinguishable images to represent the four different emotions that Google Cloud Vision API can recognize displayed in Figure 1. In the future, the app , phone, and icons will be completely discarded in favor of proprietary glasses with a discrete camera and auditory stimulation.

5. DEVELOPMENT

We used Ionic to create a smartphone app to act as a temporary stand-in for our glasses. At first, we tried to implement the sending of the image to Google Cloud API and the parsing of the response from Google Cloud API in-app. This way, we could eliminate the time needed for a middleman and maximize our response time. However, we found this to complicate things, and we could not figure out a way to properly send the request in-app. As a result, we decided to use Firebase as mentioned above. The downside to this is that now it requires Firebase to act as the middleman, so the image needs to be uploaded to Firebase, and then sent again to Google Cloud Vision API, and the response needs to come back to Firebase, be processed, and sent back to the phone.

6. EVALUATION AND RESULTS

Our app is working as intended. The benefit of using Google Cloud Vision as the engine for our emotion recognition is that is has been thoroughly trained and is constantly receiving feedback from users. However, this does not mean that it is perfect. In our small set of test data (10 images per emotion for a total of 40 images), we have noticed that Google Cloud Vision often gets confused between sorrow and anger, as well as anger and joy sometimes in some extreme cases. This is to be expected, as we also had trouble categorizing some of the images we found ourselves. In other cases, we have seen that Google Cloud Vision will not be confident in any of the four emotions even though we perceive it to be one of the four. Even though all of our communications with Google Cloud Vision have to pass through Firebase, we still managed to observe an average response time of 2.532 seconds with a standard deviation of 364 milliseconds on a strong LTE connection. When tested on a somewhat unstable WiFi connection, we observed an average response time of 2.522 seconds with a standard deviation of 1.682 seconds. Obviously, the connection strength and speed can vary and will directly affect the speed of our application.

Table 1: Response Time

Network	Avg. Response Time	Std. Dev.
LTE	2.532	0.364
WiFi	2.522	1.682

Table 2: Accuracy of Emotion Classification

Emotion	Accuracy
Anger	75%
Sorrow	60%
Surprised	85%
Joy	90%

Table 3: Most Frequent Mistakes

Correct Emotion	Incorrect Classifications
Anger	Joy
Sorrow	Anger, Joy
Surprised	Joy, Anger
Joy	Surprised

We achieved an overall accuracy rate of 77.5% on our small data set. We also purposefully chose images that could be confusing and that we had trouble distinguishing ourselves in order to see what Google Cloud Vision API is capable of. Unsurprisingly, Google Cloud Vision API confused Sorrow with Anger the most, and Surprised with Joy the most. As mentioned earlier, this makes sense, as these are emotions that could easily be misinterpreted by humans as well. What was more interesting was that Google Cloud Vision API would classify an Angry face as a Joyful face. However, in our observations, this only occurred if the person with an Angry face was exaggerating the expression and had their mouth open. This also makes sense, because the person could look excited and therefore joyful if the person had his or her mouth open. All in all, we were pretty happy with the results, and believe that they will only improve with time.

7. CONCLUSION

We found that applications of cloud computing can be successfully used in social skills therapy. In our opinion, the limiting factor of the system is the accuracy of the emotion detection itself. If emotions are incorrectly detected then it will be more difficult for the user to learn facial expressions and discern between subtle changes. Although we were unable to find statistics on the accuracy of Stanford's implementation of emotion recognition, we found that Google's implementation could be fairly inaccurate, but usually only in circumstances that had similar emotions (sorrow versus anger). We were able to achieve low response times, and further research can be done to remove the need of Firebase to improve the response time.

8. FUTURE WORK

Our next steps are to develop the hardware for the glasses, and decide an avenue for attack on the next generation of facial expression and emotion recognition. We will either work with Google to build support for more emotions in their Cloud Vision API, or we can begin developing our own infrastructure to escape Google's (albeit lofty) limitations. Perhaps we could develop a system that takes body language into account as well as facial expressions. We will also have to set up a real back-end for our wearable's front facing application that does not just route images though Google Firebase, thereby decreasing our total response time.

Acknowledgments

Thank you to Professor Soudeh Ghorbani and all of the TA's for EN.601.419 Cloud Computing for guiding us and giving feedback for this project.

9. REFERENCES

[1] Erin Digitale. Google glass helps kids with autism understand faces, stanford study finds, 2018.

[2] Matt Swider. Google glass review, 2017.

[3] Snap Inc. Shop, 2019.