Using Computer Vision and Natural Language Processing to Interpret Emotions

John Lowery
Computer Science Dept.
Florida Polytechnic University
Lakeland, Florida
jlowery9424@floridapoly.edu

***Abstract*—According to a survey taken by the National Institute of Mental Health in 2020, 21 million adults in the United States suffer from depression. While a large majority of those adults sought treatment, many either did not or could not. These numbers could also have changed since the rise of Covid-19. With in-person resources becoming less available, people are turning to technology for answers. Two such technologies are computer vision and natural language processing. These technologies can be used to provide humans a human-like interaction with a computer without having access to a professional. Artificial Intelligence and Machine Learning can use prior results from a person to give them future suggestions on possible coping methods. Python has many libraries created to allow a person to do just that. OpenCV and Microsoft Azure have features that are capable of doing all of the aforementioned tasks.**

***Index Terms*—Computer Vision, Natural Language Processing, Sentiment Analysis**

I. Introduction

A. Depression in the United States

According to the 2020 National Survey on Drug Use and Health (NSDUH), an estimated 21 million adults in the United States had at least one major depressive episode (MDE) [21]. Of these adults 66% sought out treatment. This means that 1 in 3 adults that suffered from MDE did not seek treatment in 2020. A MDE is defined by having 5 or more of the following symptoms nearly every day [21]:

1) Depressed mood most of the day
2) Markedly diminished interest or pleasure in all or almost all activities most of the day
3) Weight
4) Insomnia or hypersomnia
5) Psychomotor agitation or retardation
6) Fatigue or loss of energy
7) Feelings of worthlessness
8) Diminished ability to think or concentrate or indecisiveness
9) Recurrent thoughts of death or recurrent suicidal ideation

With the rise of Covid-19, resources became less available to people than needed them.

B. Facial Analysis Using Computer Vision

With the rise of computer vision, facial analysis has seen growth. Facial analysis is the science of using a computer to scan a person's face and analyze features from that person. Using those features we can determine characteristics from that person. For the scope of this paper we will be focusing primarily on emotional analysis. Emotional analysis allows a computer to determine a person's emotions by using a picture or a webcam. We can then use this information to gauge how a person feels.

C. Natural Language Processing

The next step for this research is to take a person's issues and cross verify with the person's emotions. To do this we can utilize natural language processing (NLP). Natural Language Processing is a form of computer science where artificial intelligence is used to process human speech. In this paper we will be using a branch of NLP called speech-to-text. This allows us to take what a person says and convert it to a form of data that a computer can process.

D. Sentiment Analysis

Another form of NLP is sentiment analysis. Sentiment analysis is taking words or sentences and detecting positive or negative connotations behind those words. This is gives us another layer of verification when determining how a person feels. Using these different branches in tandem, we can attempt to accurately determine issues a person has and give possible solutions.

II. Methodology

A. Interfacing Between Humans and Computers

To gather the data from people in a simple and friendly way, we will be using a python library called PySimpleGUI or SG for short. SG transforms the built-in python User-Interface libraries into something more developer and user friendly. We can create a simple UI to make task more understandable for a user. For demonstration purposes, the UI has two buttons. One that is used to start recording and one that closes the program. In the future buttons will be added to stop recording and to check logs from the program output. The UI will also show the user the what emotions it detects from their face and words.

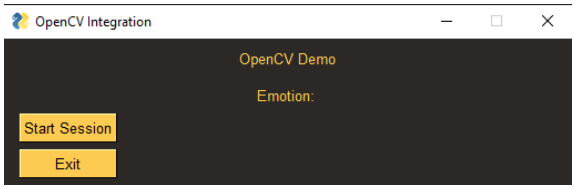

Fig. 1. What a user would see.

B. Face Capturing

After a user clicks the ”Start Session” button, a camera frame will appear on the UI. For demonstration purposes the user is not asked for permission to use their camera, but in production this will be added. A library called OpenCV is used to capture the video and seperate it into individual frames. We then use the python Threading library to constantly process the frames and update the UI. If a user wishes to stop recording, the ”Exit” button stops the recording and closes the UI.

C. Emotion Analysis

The video can then be used to preform facial analysis. To perform the emotional analysis on the user's face we use a python library called DeepFace. Deepface is a lightweight face recognition and facial attribute analysis framework for python [SO21]. It is a wrapper of other state-of-the-art facial analysis models. From DeepFace, we will be using the facial analysis features. The facial analysis features are capable of detecting: race, age, emotion, and sex. We are only concerned with the emotions for this paper so we will not be gathering data on anything else.

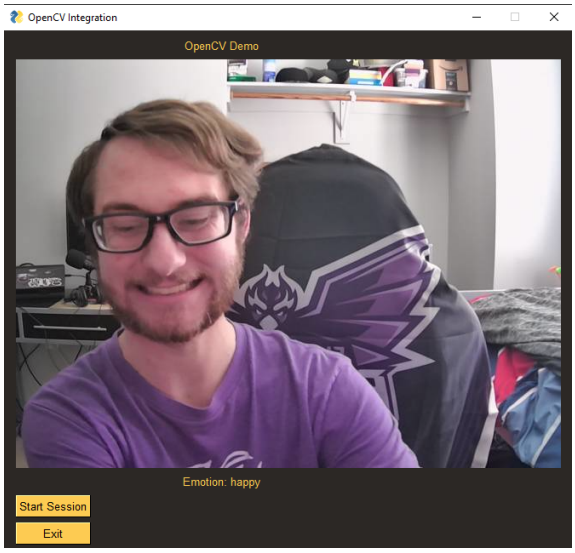

Fig. 2. The Program Detecting Emotion in Real Time

D. Speech to Text

Facial expressions are only so indicative of a real emotion, so to attempt to achieve a higher accuracy we use their speech as well. To record what a user is saying we use their default microphone and Microsoft Azure's Speech-To-Text. There are many other speech-to-text libraries but Azure was the most readily available and accurate in testing. Azure Speech-to-text allows us to take information from a user's microphone and put it in a text form that we can then process computationally. In the future, we want to choose an open-source alternative to Azure however as using it can become costly.

E. Sentiment Analysis

Using the processed data from Azure Speech-to-Text, we can then perform sentiment analysis. To obtain the connotation of the user's words we used a library called VaderSentiment. There are many sentiment analysis libraries to choose from. Many libraries are modeled using financial data however. VaderSentiment is trained using social media data. This would most closely emulate how a real person would speak. VaderSentiment gives us a score based on how positive or negative a user's sentence is. We can then cross-verify that the sentiment coincides with the emotions expressed.

F. Logging and Verification

Keeping track of output is an integral part of this project. The Logging library in python allows us to write all of the program's output to a file so we can check our results. This file is stored locally in the project folder and is overwritten every time the program runs. This file keeps track of primarily the sentiment and emotion analysis.

III. Constraints and Limitations

The first major limitation of this project is the camera. While most devices are equipped with some form of camera, not all are. The camera quality can also drastically impact the performance of the program as well. The microphone quality can also impact the results as Natural Language Processing technology is imperfect and still advancing. Microsoft Azure Speech-to-Text provides the functionality needed, it is not open-source and requires a subscription to use. The user also needs python and the libraries installed on their device to operate the program.

IV. Future Improvements

In the future we plan on moving away from Azure technologies and using libraries that are open-source. There are many alternatives but more testing is required. We also plan to package the application in a way that allows a user to download the program on a mobile device. Mobile devices are more accessible than personal computers and have cameras and microphones. The program will also ask the user for permissions to use their camera and microphone. While it is only the team using the program this is not an issue, but for the general public this could be a concern. The user interface will also be updated with more functionality. The user will be able to see the logs from the application interface. The interface will also have ”Stop Recording” buttons to allow the user to pause the recording without closing the application.

References

[21] *National Survey on Drug Use and Health 2020*. Substance Abuse and Mental Health Services Administration, 2021. URL: https://datafiles.samhsa.gov/.

[SO21] Sefik Ilkin Serengil and Alper Ozpinar. "Hyper-Extended LightFace: A Facial Attribute Analysis Framework". In: *2021 International Conference on Engineering and Emerging Technologies (ICEET)*. IEEE. 2021, pp. 1–4. DOI: 10.1109/ICEET53442.2021.9659697. URL: https://doi.org/10.1109/ICEET53442.2021.9659697.