Drawing Correlations Between Video and EEG

Andrew Festa
axf5592@rit.edu

Department of Computer Science
Department of Electrical Engineering
Rochester Institute of Technology

August 13, 2019

Abstract

The original problem attempted was that of detecting the emotional response of an individual with regards to an external, visual stimulus. By drawing correlations between how video or image affect as individual's brainwaves, it should be possible to discern their current and true frame of mind. However, this proved to be more difficult than expected and yielded no material results. Thus, the focus was shifted towards an attempt to follow an approach to use EEG responses in order to perform object segmentation. While this task was unable to be completed, the early stages of the set up and experiments proved to be more promising than those achieved with the previous task.

1 Introduction

The ability to accurately detect emotions and states of mind independently of user input could have great benefit to exploring other forms of learning or to quickly building complex datasets. For instance, a domain knowledge expert may be able to train the system in the field of social propaganda. This model could then be applied in the wild to detect potentially damaging or disparaging comments made on various forms of social media. As a cross-validation tool, it could also be used to discern between a stated response and a uncontrolled response in order to give researchers insights into why a user may provide false information to a voluntary or required question. As applied to computer vision, this type of task could help simplify steps of various algorithms which require some form of human interaction or initialization[1].

1.1 Previous Work

While much work has been done in the field of sentiment analysis, much of the research into automatic mood detection has been done by using natural language processing. EmoBank[5] provides an extensive database of several thousand, annotated sentences which span various genres and topics. WASSA[8] provides not only a database of tweets and correlated emotions, but also provides a workable model for determining the strength of a given emotion on a particular sentence. However, these are performed on textual models rather than images or videos. Even for video solutions, a common approach is to transcribe the audio and perform the analysis on the resulting text [4], [7]. Furthermore, when this type of question is posed, the emotion in question is typically that present in the video rather than of the user watching the video [3], [9].

Keeping in line with most research done in computer vision, most work regarding human interaction interfaces and EEG signals tends towards using an input image or video as the material in question[6]. This input is then directly analyzed alongside a resulting EEG signal in order to build a model of a ground-truth inherent in some form in the image. However, this task is, in some regards, both one level removed and inverse to this typical problem. In other words, rather than focusing squarely on the affect of an image on an individual, this analysis of the signal is then propagated back towards the input stimuli in order to draw further conclusions about some artefact in the input. Such an approach, similar to the one explored later, has been successfully used to automatically mark foreground and background regions of an image[2]. This then serves to initialize the GrabCut algorithm in order to perform object segmentation.

Document Version: Final Submission
Updated on: August 13, 2019

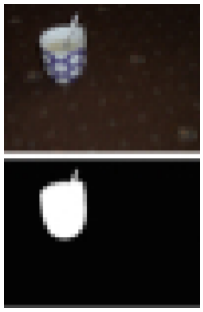

Figure 1: Example Image and Generated Mask

2 Approach

2.1 Face Recognition

The first task quickly devolved into an attempt to detect when a user recognized a face versus when they did not recognize a face. The motivation here was to build intuition surround how visual stimulus effected a change in a user's brainwaves. This first requires obtaining a set of faces unknown to the user. To this end, the face recognition database was used. More specifically, the NLPR Face Database was used. This was then augmented with 5 images of faces that the user does recognize. By presenting these images in sequence to the user without letting the user focus on each individual face, the system should be able to measure the involuntary response of a user to recognizing a face. It was believed that, in order to limit the ability of a user to focus on the image, the display rate must be about $20 fps$ due to previously seen results. However, this value was rather arbitrarily determined, and thus, the window for searching for a response to extended to 80 ms in an attempt to provide a buffer for slow reactions.

Upon measuring the resulting brainwaves given a display of a sequence of faces, it is desired to further analyze each image in order to potentially determine any artifacts in an image which may make it more difficult or simpler for the resulting model to detect if a user recognizes or does not recognize a face. This part of the approach took motivation from that high-frequency images, as well as those that are particularly composed on particular colors, would evoke a stronger response than calming scenes or colors (such as light-yellow or sea-green). Thus, each image was also measured with response to its color frequency, frequency response, and measures of energy (intensity) globally and locally.

2.2 Object Segmentation

The approach for this implementation was an attempt to replicate the results realized by another work[2]. In this paper, the authors used a set custom images, with known masks, in order to train a model to detect the foreground and background regions of an image based on a measured EEG signal as as user views sections of each image. The motivation comes from that a user will recognize a target image among various distractor images, even if they do not consciously realize as such. Thus, by mapping where a response occurs to which parts of the image are the foreground or background, the model is trained to learn what the signal looks like for a foreground pixel versus what the signal looks like for a background pixel. The paper describes this as a type of Event-Related potential and corresponds to recognizing a visual stimulus.

The paper achieved admirable results given the task. Although the resulting mask leaves much to be desired for various images or image containing multiple segmentation objects. An example image and mask is shown in figure 1. However, this is one of the better examples presented, and it is not representative of the results as a whole.

2.3 Anticipated Challenges

Any problem of interest inherently has specific parts which are more likely to be difficult than other parts. For this problem, this was anticipated to arise mainly from the EEG signals required for performing the task. While most, modern signal processing occurs at the level of tens of μV, one thing that sets this system apart is both the presence of long, thin wires along with the sensors being closely coupled to known sources of noise. The first serves to make laying out the system a challenge, as it thus requires consideration of transmission line theory in order to reduce the potential for signal reflection or cross-talk between channels. The second consideration comes into play in that hair, movement, and muscle activity can all directly influence the measured signals from the system. Thus, the most clean capture would require a user to be completely still, bald, and dead. Unfortunately, use of a cadaver presents several challenges in and of itself, not including the lack of brainwaves in such a user. As such, it is desired that the user simply limits movement as much as possible and to not blink, as this causes noticeable spikes in the signals which may be interpreted as being of meaning rather than an involuntary bodily function.

3 Experiment Setup

The system used for performing EEG measurements and data streaming was the OpenBCI Ganglion board. This board is open-hardware and is available for purchase from their site. This site also includes documentation regarding how to setup and initialize the systems necessary to perform data capture. The actual system used a Ganglion board and the code in a Python environment running in Windows 10. Normally, this would not bear mentioning. However, it is important to note that there are numerous difficulties in setting up the initial hardware to work with Windows due to the non-standard drivers Windows uses for bluetooth communication. Setup on Mac and Linux is much more direct and streamlined, but the current implementation does not support interfacing directly with the board over these operating systems.

After setting up the system, an image to be segmented must be located and placed in the local *data* directory. The next step, with regards to performing the analysis is to generate a mask of the desired segmentation. This mask is used by the algorithm not to perform the analysis, but to train and evaluate the model. The ideal situation for this algorithms is an image which has very sharp corners and whose corners are rectangular in nature, due to the method by which the distractor sections are spliced from the input image.

The third step is to actually run the image section display. This requires running the *ImagePresenter* python script, which takes in a single argument as the relative path to the desired file. This script then splits up the image into section and displays the sections in a random order with a delay of 100 ms between each section. This must be done while using the OpenBCI gui, configured during the first step and shown in figure 2, to capture the brainwaves of the user while they are viewing the image sequence of image sections.

Finally, upon completing a data capture, the GUI generates a data file (located in the *SavedData* directory of wherever the application is located). This file must be moved to the local *resources* directory, under the *eeg* subdirectory.

Upon performing these steps, the system would be able to train a model to recognize what the user's brainwave-response is to a background versus a foreground image, allowing for such a process to be gone through again in order to perform segmentation of an arbitrary image. However, as of the current implementation, only up to portions of step three are in place and able to be run.

4 Results

When running the data capture, a voltage signal, across four channels, is measured relative to a reference voltage and generates a raw data file. The heading few lines of such a data file is shown in listing 1. While accelerometer data is captured, it is not currently used. Unfortunately, no discernible responses were found

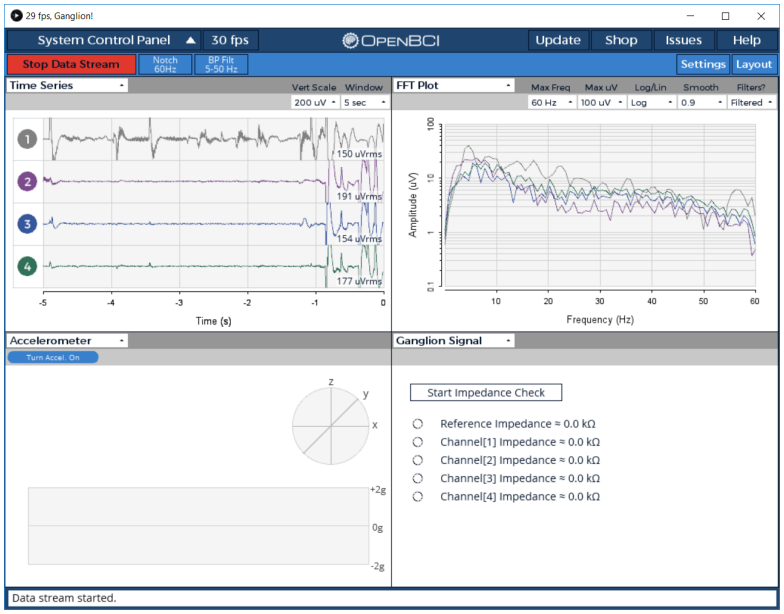

Figure 2: OpenBCI GUI

when running across a set of known versus unknown faces. Rather, the largest factor influencing the signals appeared to be a measurement of how able the user was to being able to focus on the task at hand, which could correlate to levels of sleep deprivation. Further reading suggests that the time interval over which the response was searched for in the resulting signal may have been too short. The EEG was only analyzed for 80 ms following the presentation of a known face. However, it may be possible that a response to such a stimulus does not occur for up to 300 ms following the event[2].

Listing 1: Raw EEG Data

```
%OpenBCI Raw EEG Data
%Number of channels = 4
%Sample Rate = 200.0 Hz
%First Column = SampleIndex
%Last Column = Timestamp
%Other Columns = EEG data in microvolts followed by Accel Data (in G)
%   interleaved with Aux Data
0, 0.57, 8.21, 10.34, 9.76, 0.000, 0.000, 0.000, 11:28:25.298, 1556292505298
1, 4.69, 10.71, 8.80, 8.95, 0.000, 0.000, 0.000, 11:28:25.312, 1556292505312
2, 5.29, 6.92, 4.79, 8.15, 0.000, 0.000, 0.000, 11:28:25.312, 1556292505312
3, 6.99, 16.29, 11.26, 12.90, 0.000, 0.000, 0.000, 11:28:25.328, 1556292505328
```

Pursuing the first task, that of attempting to automatically detect the true emotion evoked by an image or video, did not yield quantifiable or consistent results. However, attempting the problem required development of an interface, shown in figure 3 which is capable of reading in an image or a video and performing various transformations on the image. While the current implementation performs a fourier transform on the image and displays the resulting frequency response, other metadata regarding the image have included the magnitude and phase of the transform as well as various energy calculations at each point in the image as well as globally and locally.

For the second task attempted, the system was unable to be completed to a level whereby it could be evaluated quantitatively. However, the paper, and their described approach, highlighted and addressed some of the issues faced with the previous implementation. As such, it seems to have more promise of achieveing

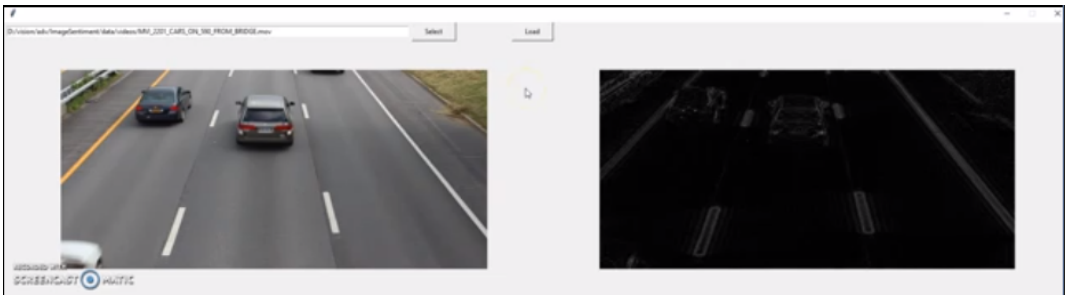

Figure 3: Image Transformation

desirable results than that of the initial approach.

5 Future Work

As of the current implementation, the EEG data must be exported in from a third-party tool, largely due to the difficulties surrounding bluetooth communication on Windows 10. However, an alternative approach was found which would use an Electron-based TCP/IP server in order to interface with the board. Another application can then interface with this hub in order to effectively stream the data from the board into the application environment in question. An implementation for interfacing with the OpenBCI hub has not been found for Python. However, one was found for $C\#$, and a next step would be to port the implementation to the existing Python code in order to speed up the process of data acquisition and pre-processing.

The second task described, that of performing object segmentation using EEG signals, still has work to be completed in order to match the implementation and results described in the paper[2]. Thus, this implementation is yet to be completed and evaluated for performance against existing methods. Furthermore, in the paper, the authors described using a Linear SVM. Thus, this could be extended simply through use of other types of classifiers, such as Naive Bayes or an RNN.

6 Conclusion

For the most part, the aforementioned projects did not yield usable results. However, this does not mean it was not without its benefits. Building the system to record and store data required finding methods to interface with the board, this requiring the development of a system to be able to either read in the raw EEG data or to directly pull data over bluetooth or wifi. Furthermore, due to the various ways in which parts of each image or signal processing failed, exploration was mainly focused on how these forms of analysis affected other tasks related to computer vision, including edge and corner detection as well as object segmentation and background extraction.

There remains much work to be done before such a proposed system could be said to be stable, and even more work before it generalizes to multiple individuals. However, these steps are the initial ones needed in order to understand the development of an end-system capable of performing deeper interfacing between computer systems and the human mind.

References

[1] Bogo. Watershed algorithm : Marker-based segmentation ii, 2016.

[2] Eva Mohedano, G. H. Object segmentation in images using eeg signals, 2014.

[3] Kotak, D. Emotion detection in videos, 2016.

[4] Krishna, A. Polarity trend analysis of public sentiment on youtube, 2018.

[5] Lab, J. Emobank, 2019.

[6] Marian Stewart Bartlett, Gwen Littlewort, I. F. J. R. M. Real time face detection and facial expression recognition: Development and applications to human computer interaction, 2003.

[7] Pérez-Rosas, V. e. a. Multimodal sentiment analysis of spanish online videos, 2013.

[8] Saif M. Mohammad, F. B.-M. Shared task on emotion intensity, 2017.

[9] Sun, Y. Authentic emotion detection in real-time video, 2004.