Towards Multiclass Brain-Computer Interface for Joint Human-Computer Image Analysis

Brent J. Lance, Anthony J. Ries, Vernon J. Lawhern, David T. Slayback, Steven M. Gutstein
Human Research and Engineering Directorate, U.S. Army Research Laboratory
Aberdeen Proving Ground, MD 21005
brent.j.lance.civ@mail.mil

Abstract

Recent advances in Brain-Computer Interface (BCI) technologies include the development of techniques that allow humans to rapidly triage images using Rapid Serial Visual Presentation, or RSVP (Sajda et al., 2010). These techniques utilize machine learning approaches to analyze electroencephalography (EEG) signals when a person sees a sparsely occurring target image. However, they are limited to target/non-target binary classification problems, and are sensitive to the ratio of target to non-target images. In contrast, most practical image classification tasks have hundreds to thousands of classes, without any guarantees to when target images will be presented. In this paper we present a novel RSVP method which is not restricted to binary classification problems, and may be more robust to target/non-target ratio, which we call the Mismatch Rapid Serial Visual Presentation (M-RSVP) paradigm. M-RSVP is based on the differences in EEG signals between viewing images with matching labels and viewing images with mismatched labels. We define the M-RSVP paradigm, present initial results obtained from a single subject, and outline plans for further research.

Introduction

There are many critical tasks that require sorting large amounts of collected image data in order to identify relatively few images of potential value that require further in—depth analysis. Ideally, this triage could be performed by an automated technology such as computer vision, which can tirelessly perform the task at a low cost relative to a human analyst. However, the performance of computer vision algorithms still does not compare to humans in many real-world image analysis tasks. Alternatively, some analysis tasks can be crowdsourced using a platform such as Amazon Mechanical Turk. Unfortunately, crowdsourcing is not a solution for all image analysis problems. Some image data, such as classified intelligence data or HIPAA-protected health data, cannot be widely distributed. Some image analysis tasks may also require rare expertise. In these cases, it is important to be able to maximize the ability of human analysts to rapidly triage images.

Background

Recent advances in Brain-Computer Interface (BCI) technologies provided human analysts with the capability to triage images more rapidly using a paradigm called Rapid Serial Visual Presentation, or RSVP (Sajda et al., 2010). RSVP is a paradigm where images from two classes, a target class and a non-target class, are presented to human subjects at a rapid rate (> 1 Hz). When the subject sees an image from the target class, an event-related potential (ERP) called the P300 can be identified from EEG data. The P300 is a large positive deflection in the scalp voltage observable approximately 300ms post stimulus presentation (McCarthy & Donchin, 1981). While early BCI systems averaged EEG data across multiple trials to identify the P300, more recent approaches have shown the capability to identify the P300 using single trials (Blankertz, Lemm, Treder, Haufe, & Müller, 2011).

While the RSVP paradigm has shown its utility as a method for image triage, it is not an ideal approach for integrating human analysts with computer vision for two reasons. First, RSVP is limited to the two-class problem of target vs. non-target images. In contrast, computer vision algorithms can provide hundreds or even thousands of potential output classes. Second, RSVP is based on an oddball paradigm requiring low probability target events (Polich & Margala, 1997). If too high a percentage of the stimuli are from the target class, the target is no longer sufficiently unusual to generate a detectable P300 response.

To help address these issues, we have developed a novel BCI paradigm called the Mismatch RSVP (M-RSVP) intended for integrating human analysts with computer vision for large-scale image analysis. In this paradigm, subjects are serially presented with images paired with a superimposed categorical label. This label either matches the contents of the image, or is a mismatched label that does not accurately describe the image (e.g., accurate/inaccurate labels provided by computer vision). By redefining the two-class problem from target/non-target images to matching/mismatched categorical labels, M-RSVP should work with an arbitrarily large number of categories while still only requiring a binary classification of neural data. The need for categorical labels for M-RSVP provides a clear mechanism for integration with computer vision algorithms by using the computer vision to preselect a small subset of labels for each image.

In addition, M-RSVP is based on Stroop-style image/word interference paradigms (see Bajo, 1988; Glaser & Düngelhoff, 1984; Pellegrino et al., 1977; Snodgrass & McCullough, 1986). Unlike the Oddball paradigm that the standard RSVP task is based on, Stroop interference tasks are less dependent on low-probability target events. As a result, M-RSVP may be less sensitive to target probability ratios. In this paper, we will define M-RSVP, present preliminary results obtained from a single subject, and outline plans for further research in this area.

Methods

During M-RSVP, we provide subjects a rapid presentation of a series of stimuli consisting of individual images. A brief interval (0.25-1s) after each image is displayed, a paired categorical label is superimposed onto the image. After another brief interval the process repeats. Images and categorical labels from the stimuli were drawn from the Places2-365 image database, which consists of 365 categorical labels of locations, with 5,000-30,000 example images for each categorical label (Zhou, Khosla, Lapedriza, Torralba, & Oliva, 2016). A stimulus can either be a matching stimulus, where the paired categorical label accurately identifies the contents of the image (see Figure 1), or a mismatched stimulus where the paired categorical label does not accurately describe the contents of the image (see Figure 2). We used a standard counting procedure, where subjects silently counted the number of mismatched stimuli that occurred in an experimental block.

Experimental Details

One subject (Male, 38 years old) performed ten 2-minute blocks of the M-RSVP paradigm. Stimuli, consisting of a single image and the corresponding label were presented sequentially for one second each, resulting in an overall presentation rate of 0.5Hz. This slow rate was chosen to minimize the overlap of the ERP signals but early pilot testing indicates that considerably more rapid presentation rates are possible. To further reduce ERP overlap, both image and category label onset were randomly jittered by up to ±100ms. Eighty percent of the stimuli presented were matching stimuli, with the remaining 20% being mismatched stimuli.

EEG data were recorded at 256Hz using a Biosemi ActiveTwo system with 64 active Ag/AgCl electrodes. Vertical (VEOG) and horizontal (HEOG) electrooculogram were recorded with VEOG electrodes centered superior and inferior to the left eye and HEOG electrodes placed along the outer canthus of each eye. In addition, reference electrodes were placed on both the left and right mastoid.

Figure 1: Matching M-RSVP Stimulus. A brief interval after each image is displayed, a categorical label that describes the image is superimposed on top of the image.

Figure 2: Mismatched M-RSVP Stimulus. A brief interval after each image is displayed, a categorical label that inaccurately describes the image is superimposed on top of the image.

Analysis

EEG data were re-referenced to averaged mastoids, high pass filtered at 0.1Hz and low pass filtered at 40Hz. The EEG data stream was then epoched into 1-second epochs timelocked to the onset of category label events. Averaged ERPs were generated for both the matching stimuli and mismatched stimuli, and an ERP showing the difference between these two ERPs was generated.

Results

The resulting ERPs can be seen in Figure 3. There appears to be a separation between the matching and mismatched stimuli starting at approximately 400ms after category label presentation.

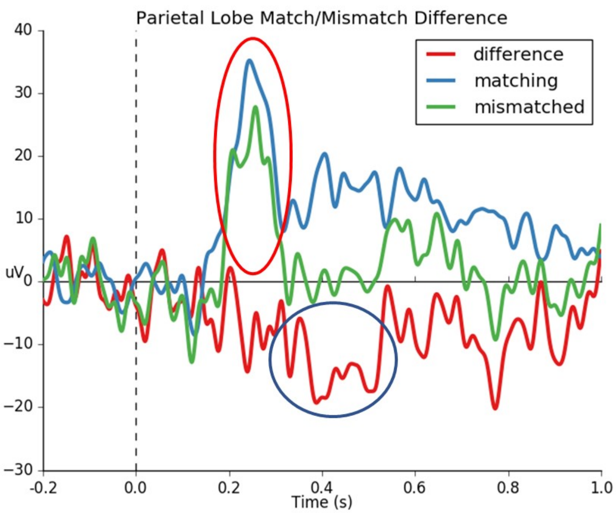

Figure 3: Averaged neural responses resulting from the M-RSVP paradigm. The blue ERP is the averaged neural response of the matching stimuli. The green ERP is the averaged neural response of the mismatched stimuli. The red ERP is the difference between the two. A red oval highlights a P300 response, and a blue oval highlights a likely N400 response in the red difference wave.

Discussion

While these results are preliminary, they are also very promising. First, an N400 response can be clearly seen in the neural data. The N400 response is characterized by a negative deflection in the voltage recorded at the scalp occurring 400ms after stimulus presentation. The N400 is evoked through semantic discrepancies in image or text stimuli (Kutas & Federmeier, 2011). While the N400 response has seen much less interest in the BCI community than the P300 due to its much smaller amplitude (see Figure 3), there have recently been BCI paradigms developed that incorporate the N400 response (Jin, Allison, Zhang, Wang, & Cichocki, 2014; Van Vliet, Mühl, Reuderink, & Poel, 2010). There have also been BCI paradigms developed using error-related potentials, which are similar in amplitude to the N400 (Barachant & Congedo, 2014).

Second, while the data in this paper was collected at a very slow presentation rate (0.5 Hz), this was done to maximize separation of ERPs. Preliminary testing indicates that it may be possible to run the M-RSVP at a rate of up to 2Hz. We note that M-RSVP will likely show higher levels of performance when the semantic distinctions between the categorical labels are broad. However, this may not be a major limitation, as previous work shows that computer vision provides higher performance than humans at fine-grained recognition in domains with large numbers of categorical labels (Russakovsky et al., 2015).

Future Work

The immediate next step is to execute a full human-subject study that better validates the M-RSVP paradigm, and define the parameter space in which it is effective. After that, it will be necessary to test the M-RSVP paradigm as a closed-loop BCI with real time classification and feedback.

Conclusion

In this paper we introduced a novel BCI paradigm, which may have value for joint human and computer vision analysis of large image databases. While the results are preliminary, they have several promising implications in joint human-computer image analytics. For example, M-RSVP could potentially be used in conjunction with Active Learning to iteratively update a computer vision system. In the future, we intend to integrate a paradigm such as this into a human-computer image analysis system such as the Cortically-Coupled Computer Vision system or the Human-AI Image Labeling system (Saproo et al., 2016).

Acknowledgements

This project was sponsored by the U.S. Army Research Laboratory under CAST 076910227001 and ARL-H70-HR52.

References

Bajo, M.-T. (1988). Semantic facilitation with pictures and words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 579–589. https://doi.org/10.1037/0278-7393.14.4.579

Barachant, A., & Congedo, M. (2014). A plug & play P300 BCI using information geometry. Retrieved from http://arxiv.org/abs/1409.0107

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K.-R. (2011). Single-trial analysis and classification of ERP components--a tutorial. *NeuroImage*, *56*(2), 814–825. https://doi.org/10.1016/j.neuroimage.2010.06.048

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 640–654. https://doi.org/10.1037/0096-1523.10.5.640

Jin, J., Allison, B. Z., Zhang, Y., Wang, X., & Cichocki, A. (2014). An ERP-Based BCI using an Oddball paradigm with different faces and reduced errors in critical functions. *International Journal of Neural Systems*, *24*(8), 1450027. https://doi.org/10.1142/S0129065714500270

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123.Thirty

McCarthy, G., & Donchin, E. (1981). A Metric for Thought: A Comparision of P300 Latency and Reaction Time. *Science 211.4477*, 77–80. https://doi.org/10.1126/science.7444452

Pellegrino, J. W., Rosinski, R. R., Chiesi, H. L., & Siegel, A. (1977). Picture-word differences in decision latency: An analysis of single and dual memory models. *Memory & Cognition*, *5*(4), 383–96. https://doi.org/10.3758/BF03197377

Polich, J., & Margala, C. (1997). P300 and probability: Comparison of oddball and single-stimulus paradigms. *International Journal of Psychophysiology*, *25*(2), 169–176. https://doi.org/10.1016/S0167-8760(96)00742-8

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., … Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Sajda, P., Pohlmeyer, E., Wang, J., Parra, L. C., Christoforou, C., Dmochowski, J., … Chang, S.-F. (2010). In a Blink of an Eye and a Switch of a Transistor: Cortically Coupled Computer Vision. *Proceedings of the IEEE*.

Saproo, S., Faller, J., Shih, V., Sajda, P., Waytowich, N. R., Bohannon, A., … Jangraw, D. (2016). Cortically Coupled Computing: A New Paradigm for Synergistic Human-Machine Interaction. *Computer*, *49*(9), 60–68. https://doi.org/10.1109/MC.2016.294

Snodgrass, J. G., & McCullough, B. (1986). The role of visual similarity in picture categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(1), 147–154. https://doi.org/10.1037/0278-7393.12.1.147

Van Vliet, M., Mühl, C., Reuderink, B., & Poel, M. (2010). Guessing what's on your mind: Using the N400 in brain computer interfaces. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 6334 LNAI, pp. 180–191). https://doi.org/10.1007/978-3-642-15314-3_17

Zhou, B., Khosla, A., Lapedriza, A., Torralba, A., & Oliva, A. (2016). Places: An Image Database for Deep Scene Understanding. *arXiv:1610.02055 [Cs]*. Retrieved from http://arxiv.org/abs/1610.02055