Neurophysiologically Driven Image Triage: A Pilot Study

Abstract

Effective analysis of complex imagery is a critical aspect of important domains such as medical diagnosis. As technological developments lower the cost of gathering and storing imagery, the cost of searching through large image sets for important information has been growing substantially. This paper demonstrates the feasibility of using neurophysiological signals associated with early perceptual processing to identify critical information within large image sets efficiently. Brain signals called evoked response potentials, detected in conjunction with rapid serial presentation of images, show promise as a human computer interaction modality for screening high volumes of imagery accurately and efficiently.

Keywords

Rapid Serial Visual Presentation, EEG, Visual Search, Brain Computer Interface

ACM Classification Keywords

H.5.2. Information interfaces and Presentation: User Interfaces: Interaction Styles H.3.3 Information Search and Retrieval: Information Filtering

CHI 2006, April 22–27, 2006, Montreal, Canada.

ACM 1-xxxxxxxxxxxxxxxxxx.

Introduction

The problem of searching for targets in vast collections of imagery is one that affects practitioners in a variety of

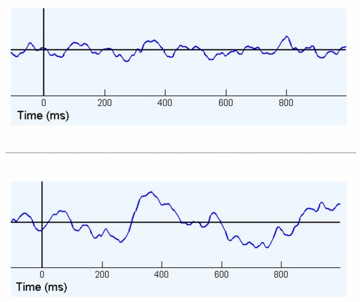

Figure 1. Baseline EEG (top) EEG segment containing an Evoked Response Potential (bottom)

domains – from medical diagnosis to intelligence image analysis. Advances in imaging and storage technology have served to lower the cost of collecting and storing high volumes of imagery. However, the cost of searching through large sets of imagery for important information can often be substantial. In many domains, such as medical diagnosis and intelligence analysis, effective search currently requires the expertise of highly skilled analysts who search though sequences of images in a relatively slow manner. Unfortunately, the availability of skilled analysts is simply insufficient to cope with the volume of imagery to be analyzed. For example, the military reports that most intelligence imagery goes without being examined by analysts [4].

The problems just highlighted have led to calls for effective triage techniques that can be used to rapidly screen high volumes of imagery and identify a subset of images that merit careful scrutiny by an image analyst [4]. A triage process trades off specificity in favor of sensitivity – the search may result in several false positives, but also most, or all, of the targets in an image set. Computer vision systems have been employed towards this end. However, in many contexts, these systems fall short of the sensitivity and specificity that humans display. They also fail to generalize to the extent that human analysts do. An ideal triage system might be one that leverages human visual processing capabilities in the role of a target detector, while raising the efficiency associated with the manual search process.

One method for realizing an efficient triage platform that exploits human perceptual capabilities may lie in a combination of rapid serial visual presentation (RSVP) of images and electroencephalogram (EEG) recordings collected as a user views images flashed briefly at rates of 100 milliseconds or less per image. For example, Thorpe and colleagues [8] asked participants to detect images of animals in a sequence of nature scenes presented for 20 milliseconds per image. Using EEG sensors, researches were able to detect a brain signal known as an *evoked response potential* (ERP) within 150 ms of the onset of target stimuli.

Evoked Response Potentials

Evoked response potentials refer to a morphological change in EEG waveforms in response to task-relevant stimuli. They are typically measured by inspecting EEG activity within a window of several hundred milliseconds following critical events. Figure 1 shows EEG activity at a particular sensor following a non task relevant stimulus (distractor) and a task relevant stimulus (target). The x-axis depicts the progression of time following the stimulus in milliseconds — the zero point corresponds to the onset of a stimulus. The wave form associated with the target shows a pronounced amplitude perturbation following stimulus onset.

Research suggests that ERPs reflect the activity of underlying cognitive processes necessary for processing and coordinating a response to task relevant stimuli. The brain’s response to critical events, such as the presence of targets, may begin in frontal areas — generating top-down, intent information — and propagate to sensorimotor areas – triggering events that regulate bottom up information transmission through sensory and response selection areas. [5]

ERPs are difficult to detect. These signals typically range in amplitude from approximately 1 to 10 microvolts, while background EEG activity may range from 10 to 100 microvolts. Common events such as eye blinks or facial

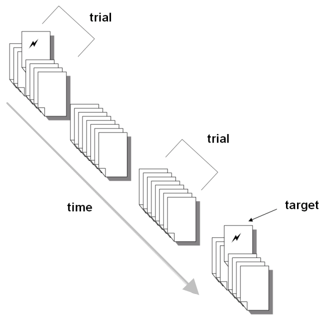

Figure 2. Experimental design. Participants viewed trials with or without targets. 50% of trial blocks contained targets. Fixation screen separated trial blocks.

muscle activity can completely obscure ERPs. In order to deal with such an inherently low signal to noise ratio, ERP detection has relied on a strategy of trial averaging. Under this strategy, an experimental stimulus is presented to a subject several times. The waveforms elicited by each stimulus are averaged. Background EEG washes out in the averaging process, and the event induced activity becomes prominent.

While integrating information across repeated presentations of a stimulus is an effective way to identify ERPs, it is an impractical strategy for application domains, such as a triage platform. Repeated presentation of stimuli compromises the efficiency of the search process. In domains where efficient ERP detection is critical, accurate detection of ERPs within a single trial becomes necessary. However, single trail detection of ERPs requires a robust signal processing and classification approach to overcome the problems imposed by the inherently low signal-to-noise ratio.

Recently, researchers have developed promising approaches for single-trial ERP detection [3, 6]. Instead of integrating sensor data over time, they rely on integrating information spatially, across high density EEG sensors. While these studies are promising, several issues limit their practical relevance. First, they rely on large electrode arrays of over 60 elements — this is both expensive and cumbersome. Second, the ability of classifiers to generalize across individuals and sessions spanning an hour or more has not been explored. Research to date has largely focused on data collected over the span of sessions under 10 minutes. However, analysts anecdotally report analyzing imagery for spans of approximately an hour. Third, existing research has focused on within-subject analysis of classification results. However, the ability of an ERP detector to generalize across subjects has practical implications. An approach that generalizes across individuals would reduce the need for individualized calibration and allow a greater proportion of each analysis session to be devoted to image analysis.

The pilot study described in this paper focuses on ERP classification in the context of an RSVP based triage task. Using data collected with a 32 electrode EEG system, this research examines the capacity of an ERP detection system to generalize over the span of an hour. It also examines the ability of the system to generalize across individuals.

Method

Participants

Two male volunteers participated in this pilot study. Both had normal corrected vision. EEG data was collected over the course of little over an hour as participants performed a target detection task in the context of an RSVP presentation. The experimental duration was broken into three sessions of approximately 20 minutes and produced three corresponding datasets for each subject. There was a rest period of approximately five minutes between sessions. Data from the first session was intended for use as training data for a classifier, data from the second session was intended for classifier validation, and the third dataset was collected with the objective of being used for classifier testing

Task

Participants were asked to locate objects within sequences of grayscale satellite imagery provided by the National Geospatial Intelligence Agency (NGA). Objects of interest, referred to as targets, consisted of satellite photographs of ships or boats in the midst of a pool of satellite images around a port scene. Both the target and distractor images

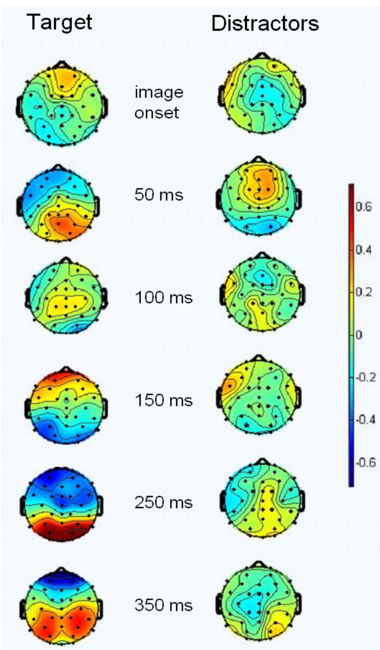

Figure 3. Average spatio temporal pattern of electrical activity over the scalp following target (left) and distractor images (right)

were drawn from a common high-resolution, broad-area, satellite image. All imagery was presented using the Rapid Serial Visual Presentation (RSVP) paradigm [7]. Images were presented in rapid succession for durations of 50 or 100 milliseconds, per image. These images were grouped into sets of approximately 50 images, referred to as a trial (see Figure 2). Half the trials included a single target image inserted randomly within the sequence of images (target trials), while half the trials did not include a target (distractor trials). Participants were asked to indicate the presence of a target with a button press. Consecutive trials were separated by a fixation screen lasting several seconds in order to break monotony and minimize possible eye strain. Participants viewed tens of thousands of images over the course of the experiment.

Display

Images were presented on a 21 inch, CRT monitor. Images were 400 x 400 pixels in size and presented on a screen of 1240 x 768 pixel resolution. Participants were able to position themselves at a comfortable distance from the screen. All images shared a relatively similar level of luminance and were presented using a script developed for Presentation, a stimulus presentation tool developed by Neurobehavioral Systems.

Data Acquisition

EEG data was collected using the BioSemi Active Two system using a 32 channel EEG cap and eye electrodes. Data was sampled at 256 Hz. Triggers sent by the Presentation script to mark the onset of target and distractor stimuli were received by the BioSemi system over a parallel port and recorded concurrently with EEG signals. EEG was bandpass filtered between 1 Hz and 30 Hz, using an 8th order Butterworth filter. An adaptive linear filter was used to correct EEG signals affected by eye blinks.

Results

Following the experiment, EEG data was segmented into epochs. In the case of target trials each epoch consisted of a two second segment of EEG — one second before, and one second after the onset of target stimuli. For distractor trials (no target trials) epochs were extracted around the trigger associated with the middle image of each trial block. Each epoch served to provide a picture of spatio-temporal electrical activity across brain regions. Each twenty minute session yielded approximately 80 to 90 target and 80 to 90 distractor epochs each. Each independent channel of each epoch (indexed by row in the epoch matrix) was normalized using the mean and standard deviation of channel values preceding the image trigger.

Qualitative Analysis

MEAN HEAD PLOTS

Epoch associated with distractor and target images were averaged and rendered using EEGLab [1]. The visual rendering of spatio-temporal electrical activity reveals clear patterns of activation that help discriminate between target and distractor trials for both participants. For example, as Figure 3 based on data from Participant 1 shows, target trials were characterized by strong, positive, frontal activity and negative parietal activity at approximately 150 milliseconds. This pattern changes to strong negative frontal activity and positive parietal activation 350 milliseconds after stimulus onset. In contrast, activations remain relatively neutral and stable in the distractor conditions. Clear patterns of discriminatory activity are evident well beyond the 450 millisecond window depicted in Figure 3.

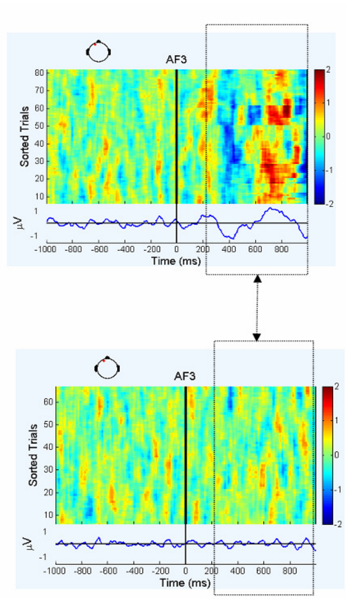

Figure 4. Electrical activity preceding and following target images (top) and distractor images (bottom) at a single EEG electrode

Y-axis in each plot indexes individual trials or epochs; the x-axis represents time. The 0 point on the x-axis (bold black line) represents time of stimulus onset. Color in each graph represents the polarity and amplitude of electrical activity at a particular point in time.

There is a consistent pattern of sustained, high amplitude activity following stimulus onset at practically every trial containing a target.

While the mean head plots depicted in Figure 3 reveal a clear spatio-temporal pattern of activity that could serve to discriminate between targets and distractors, they provide no information about the trial-by-trial variability of activity patterns across time at particular channels. A high degree of variability could make it difficult to discriminate between targets and distractors on a trial-by-trial basis. To assess the variability of patterns that form features for classification, electrical activity associated with every trial was plotted for several channels. Figure 4 shows a channel ERP image for Participant 1 plotted using EEGLab. The channel ERP image depicted in Figure 4 shows electrical activity associated with every epoch in target and distractor conditions, at a particular frontal sensor location. Target epochs contained broad bands of high amplitude activity at various latencies with respect to stimulus onset. By comparison, sustained patterns of high amplitude activity are missing in the distractor epochs.

ERP Classification

Objective

ERP data was also used in the context of a classification analysis. There were two objectives associated with the analysis. First, the analysis would serve to establish whether it was possible to generalize learning over relatively large time windows — up the span of an hour. Second, the analysis would help assess whether the features used to classify ERPs were consistent enough to generalize across individuals.

A support vector machine (SVM) [2] was used in conjunction with this effort. Support vector machines are a widely-used linear machine learning technique that relies on ideas from statistical learning theory to provide good generalization performance. Support vector machines can also be used in the context of problems that are not linearly seperable by projecting data into a higher dimensional space where the data may be linearly seperable. A non-linear support vector machine with a radial basis function kernel was used in this study.

The metric used to evaluate classification performance is the Area under the Receiver Operating Characteristic (ROC) curve [2]. ROC curves plot the ratio of True Positives to False Positives as a threshold for discriminating between targets and distractors is varied. It is widely used to evaluate human and machine signal detection capabilities. Perfect classification produces an Area under the curve value (Az) of 1.0, while chance performance produces an Az value of 0.5.

Generalization over time, within subject

As mentioned earlier, many single trial ERP classification studies have relied on a leave-one-out testing procedure to assess their classification approach. Such an approach largely circumvents well known problems associated with changes in the statistical properties of biological signals over time. In contrast, this study examined the capacity of a classifier to generalize over experimental sessions that span over an hour.

A classifier trained on data from the first 20 minute session generalized well to data from a 20 minute test session separated by over a twenty minute gap. For Participant 1, an SVM produced an area under the receiver operating curve (Az) of 0.90. Results for Participant 2 were similar — an SVM produced an Az value of 0.91. These results suggest that the features that help discriminate between EEG in epochs with target and distractors remain stable over the course of sessions distributed over the span of an hour.

Generalization across individuals

In order to test the ability of the ERP detection approach described here to generalize across participants both the training and test data for Participant 1 were used as training data for the classifier. Both the training and testing data for Participant 2 were combined into a test data set. An SVM trained and tested with this data produced an Az value of 0.84. Similar results were obtained when the training and test dataset were reversed(training and test data for Participant 2 were aggregated into a training set, while training and test data for Participant 1 were aggregated into a test set). An Az value of 0.85 was obtained when training and test data were reversed.

These results suggest that there is a stable and similar pattern of spatio temporal electrical activity induced by targets and distractors in both individuals. Whether these cross-subject classification results are sufficient for a triage platform will depend on the application domain and the cost of misclassification.

Discussion

The results presented here suggest that evoked response potentials detected in conjunction with rapid serial visual presentation of images offer a viable interaction modality for searching through complex imagery efficiently and accurately. Classification results suggest that features that serve to discriminate among conditions remain stable over time and may be shared among individuals. Additionally, the results reported here suggest that an array of 32 EEG electrodes may provide a sufficient set of features for detecting ERPs with a high degree of accuracy. Future work will assess the efficacy of the image triage approach described here across a wider group of participants.

Acknowledgements

This research was funded by the Defense Advanced Research Projects Agency (DARPA) and the National Geospatial Intelligence Agency (NGA). This paper has been reviewed by DARPA and NGA and Approved for Public Release, Distribution Unlimited.

References

[1] Delorme A and Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics J Neuroscience Methods, 134:9-21, 2004

[2] Duda, R.O, Hart, R. E. and Stork D. G. Pattern Classification, Second Edition. John Wiley & Sons, New York, 2001

[3] Gerson, A.D., Parra, L.C., and Sajda, P. (2005) Cortical Origins of Response Time Variability during Rapid Discrimination of Visual Objects, NeuroImage, 28 (2) 326-341.

[4] Kenyon, H. S. (2003). Unconventional Information Operations Shorten Wars. Signal Magazine. Armed Forces Communications and Electronics Association.

[5] Makeig S, Westerfield M, Jung T-P, Enghoff S, Townsend J (2002) Dynamic brain sources of visual evoked responses. Science 295: 690–693

[6] Parra, L., Alvino, C., Tang, A., Pearlmutter, B., Yeung, N., Osman, A., and Sajda, P. (2003), Single Trial Detection in EEG and MEG: Keeping it Linear. Neurocomputing. (52-54) 177-183.

[7] Spence, R., Witkowski, M., Craft, B., de Bruijn, O. & Fawcett, C. (2004) International Conference Advanced Visual Interfaces (AVI-2004). pp. 141-149

[8] Thorpe,S., Fize, D. & Marlot, C.(1996). Speed of processing in the human visual system. Nature, 381, 520-522.