

# LARGE-SCALE IMAGE DATABASE TRIAGE VIA EEG EVOKED RESPONSES

Yonghong Huang<sup>1</sup>, Deniz Erdogmus<sup>1</sup>, Santosh Mathan<sup>2</sup> and Misha Pavel<sup>1</sup>

<sup>1</sup> Oregon Health & Science University, Portland, OR, USA

<sup>2</sup> Human Centered Systems Laboratory, Honeywell, Redmond, WA, USA

## ABSTRACT

This paper describes an approach for target image search using human brain signals generated by perceptual processes in the brain. The human brain generates event related potentials (ERPs) in response to critical events, such as interesting/novel visual stimuli in the form of a target image. In this paper, we describe experiments involving six professional image analysts and summarize the ERP detection performance as they search for targets within a large image database. We develop a disjoint windowing scheme for data preprocessing to discard irrelevant and redundant information from the raw data to get clean training data. We apply support vector machines to detect ERPs and conduct 10-fold cross validation for parameter regularization. The results demonstrate that the ERP pattern recognition can provide reliable inference for image triage.

*Index Terms*— event related potentials, EEG, image triage, brain computer interface, pattern recognition

## 1. INTRODUCTION

Image search through large volumes of images has become an important issue in many domains. In search tasks, human experts display great skills in exploiting contextual cues and prior knowledge to deal with variability within and across images. In contrast, fully automated target detection algorithms are still not feasible. Recently researchers began to exploit signals associated with split second perceptual judgments as the basis for image triage [1-5].

Our solution to the problem of target image search in a vast database is to develop an image triage platform to rapidly process the images and identify a subset that deserves careful inspection by a human expert. The triage process is driven by exploiting the visual and cognitive systems of the human expert. The goal of our research is to develop an effective triage platform to increase the efficiency of image search. The system exploits electroencephalography (EEG) as the main indicator to see if an image seen briefly by the expert contains a target (object of interest) or a non-target (distractor). The main task is to detect the event-related potentials (ERPs) corresponding to target stimuli. Our previous works [6-8] demonstrate our ERP-based image triage system is viable for target image search. Here we describe a disjoint windowing scheme to extract EEG data and apply a support vector machine (SVM) as the ERP detector. The results

establish that the system based on brain signal monitoring is capable of detecting targets from a large image set efficiently.

## 2. ERP AND IMAGE TRIAGE

The ERP-based image triage system collects and analyzes EEG signals by monitoring brain activity as a subject performs a high speed scan of large image sets. Figure 1 illustrates the structure of the system.

### 2.1 Image Display – RSVP Modality

We adopt a rapid serial visual presentation (RSVP) protocol for image presentation. The work of Thorpe et al. has demonstrated that the ERP signal can be used as a target detection cue within a large image set [3]. As shown in Figure 2, during the RSVP search, a sequence of images is rapidly presented. A target image in a sequence of distracter images elicits an ERP, a pronounced amplitude perturbation in the EEG waveforms.

### 2.2 ERP vs. Non-ERP

The core task is to apply pattern recognition techniques to detect ERP patterns. An ERP is a stereotypical electrophysiological response to a stimulus. Recent research has demonstrated that ERPs can reveal signs of neural processing well before motor outputs [2]. Figure 3 is the ERP vs. non-ERP image plots corresponding to targets and distractors. One can observe a clear ERP pattern corresponding to targets while no pattern to distractors. There is also a perturbation in the bottom trace associated to target stimuli. The main challenge of the ERP detection is low signal-to-noise ratio of an ERP (Background EEG can be 10 fold higher than an ERP). Eye blinks or facial muscle movement may smear the ERP signals. The conventional strategy for the ERP detection is averaging across trials [1,3]. However, the trial-averaging compromises the efficiency of image search and thus is infeasible for a triage platform. Parra and colleagues developed a promising approach for single-trial ERP detection [9]. Instead of integrating sensor data over time, the spatial information across EEG sensors was integrated.

### 2.3 Data Collection

Six professional image analysts (IAs) were recruited for the study. None of them had experience with the RSVP modality. The broad-area aerial images were decomposed

into hundreds of smaller chips (500x500) and were labeled whether or not they contained targets. The chips were presented at very high rates (durations were 60ms, 100 ms, or 150ms per chip) in the RSVP paradigm. The subjects performed target detection on an RSVP task (Fig. 2) by clicking on a button as soon as they saw a target. At the same time, we monitored their brain signals (EEG) and stored the data for subsequent analysis.

We used two computers to acquire data, one for image display and one for data collection. The EEG data were collected using a 32-channel Biosemi ActiveTwo system. Presentation™ (Neurobehavioral Systems, Albany, CA) software was used to present images with a high degree of temporal precision and to output pulses or triggers to mark the onset of target and distractor stimuli. The triggers were received by the Biosemi system over a parallel port and recorded concurrently with the EEG signals. The user's button presses of indicating the response to target presence were recorded by the Biosemi system as well.

### 2.4 Image Triage

Our goal is to develop an effective image triage system to leverage expert human resources. Using this technique, a human expert is able to rapidly screen high volume of images, based on which the system can assign the priority order to the images and identify a subset of images that deserve careful inspection. The system sorts the images by the estimated likelihood of each image being a target and selects a subset of the images with the highest likelihood values.

## 3. ERP-DETECTOR CONSTRUCTION

### 3.1 Data Pre-Processing

The data comprised six training sessions and 33 test sessions. Each subject had one training session and several test sessions. The duration of each image for most subjects was 100ms except subject #3 and #5. For subject #3, the image duration in the training session and four test-sessions was 60ms; the image duration of three test-sessions was 100ms. For subject #5, the image duration of the training and test was 150ms. In the training sessions the images were randomly drawn from the image chip set while in the test sessions the image chips were displayed in the natural spatial order of the broad-area image. We randomly positioned around 50 target images in hundreds of distractor images during the training while only one to eight target images within thousands of distractor images during the test. Therefore the data are very unbalanced. During the test for subject #3, #4, #5 and #6, *fake targets* that were not part of the original broad-area image were introduced randomly to keep the subject alert to prevent the boredom-related ERP-degeneration. We excluded these fake targets when conducting the performance evaluation of target detection.

We pre-processed the EEG data to extract the most relevant information from the raw EEG signals using

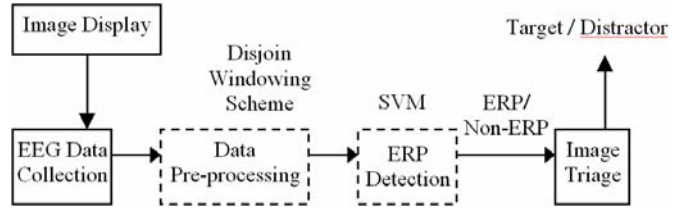


Fig.1: The ERP-based image triage system

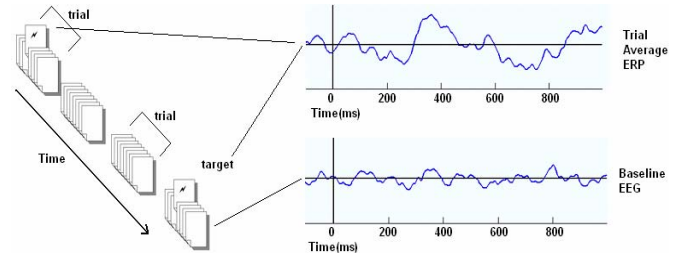


Fig. 2: The RSVP image display modality. The upper trace is an ERP averaged over trials. The lower trace is the baseline EEG signals. The zero point of x-axis corresponds to the stimulus onset.

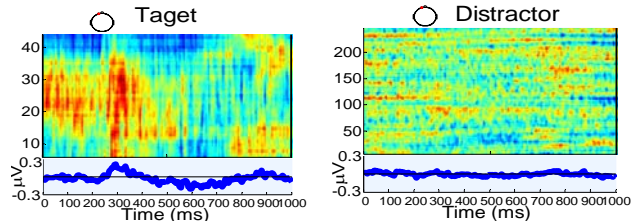


Fig. 3: The ERP images for subject #1 at channel FP1. The images show electrical activity following target and distractor images. The y-axis is trial number. The bottom traces are the EEG signals averaged over trials. The zero point corresponds to the stimulus onset.

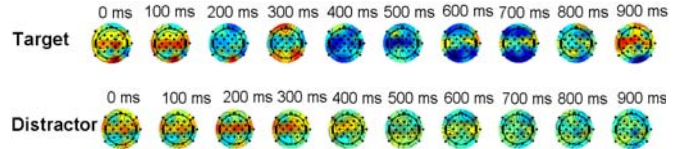


Fig. 4: The ERP scalp images for subject #1 on channel FP1. The images show average spatiotemporal pattern of electrical activity over the scalp following target (top) and distractor images (bottom).

several procedures. First, we segmented the data into the task-relevant epochs. Each epoch consisted of a short segment of EEG (approximately 500 ms after each image trigger). The raw EEG data were used without filtering. Figure 4 shows the electrical activity over the scalp over time from the image onset to 900ms. One can see that there is a peak around 300ms for target stimuli while no magnitude change exists for distractor stimuli. There are some magnitude changes after 600ms, which are mainly due to motor responses (button clicks). To extract the most relevant portion of the ERP, we truncated each epoch in the interval of the stimulus onset to 500ms. Each epoch represented the spatiotemporal electrical activity across brain regions associated to a stimulus. We used the 100 ms

portion preceding the stimulus onset to normalize the data, and rescaled the data to  $[0, 1]$ . Second, we adopted a disjoint windowing scheme in the training sessions and a sequential windowing scheme in the test sessions. The goal of disjoint windowing scheme was to provide clean ERP signals for classifier training. The disjoint windowing scheme reduced the data dimension and provided a more balanced (targets vs. distractors) training data. Figure 5 shows the disjoint windowing scheme. Each disjoint window size is 600ms, 100ms before the trigger (normalized window) and 500ms after the trigger (epoch window). We only extracted the disjoint windows of EEG data and discarded the data overlapping within each window. Third, for both training and test sessions, we removed the distractors in the interval of one second before and after the targets to eliminate the overlapping information in the samples. Four, for both training and test sessions, we removed the targets without button clicks and only selected the target with following button responses within 1.5 seconds for both training and test sessions. We assumed that there was no ERP if there was no button clicks following the targets. If there was more than one target in sequence, we only selected the first target. The 32-channel data in each epoch were eventually congregated to form a feature vector and the raw EEG measurements were subjected to the classifier.

### 3.2 ERP Detector - Support Vector Machine

Our goal in classification is to build an ERP detector to accurately detect the ERPs associated with target stimuli. We adopt SVM [10,11] as an ERP detector. A radial basis (Gaussian kernel) SVM is used in this study. The SVM is optimized to construct a maximum-margin separating hyperplane by mapping input vectors to a higher dimensional space. The separating hyperplane is the hyperplane that maximizes the distance between the two parallel hyperplanes on each side of the boundary touching closest data (support vectors) from each class. The assumption is that the larger the margin between these parallel hyperplanes the less the generalisation error will be. A cost parameter  $C$  in the optimality criterion controls the number of support vectors and the trade-off between learning error (margin) and model complexity (the size of the slack variables). A larger  $C$  corresponds to assigning a higher penalty to errors (when the classes are not separable by a hyperplane in the feature space). To find the optimal hyperplane, the SVM is trained and optimized by solving a convex quadratic programming problem. After training, the optimal Lagrange multipliers for each sample and weights are obtained. Support vectors, which are the data points lying at the border of the margin have non-zero optimal solutions for their coefficients in the final discriminant, while others converge to zero weights, thus leading to a sparse nonparametric forward discriminant function.

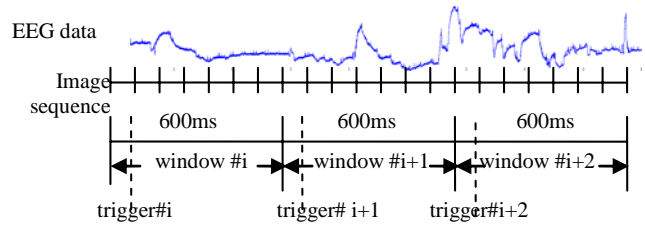


Fig.5: Disjoint windowing scheme of continuous EEG data. Each disjoint window is 600ms, 100ms before the trigger and 500ms after the trigger.

The kernel size  $k$  and the cost parameter  $C$  can be chosen by users. To avoid overfitting, we adopt 10-fold cross-validation [11] to adjust the regularization parameters. The cross-validation procedure is conducted on the training session to select the optimal parameters and the parameters are then applied to the *independent* test data (collected in a session immediately following the training session) to do the classification.

### 3.3 Evaluation Criteria and Parameter Regularization

Due to the probability of targets being extremely low, we aim for a minimum false alarm rate for zero-miss (MFAR) strategy. Define the *MFAR* as  $N_m/N_D$ , where  $N_m$  is the number of distractors for which discriminant values are higher than the minimum discriminant values of targets in the test set and  $N_D$  is the total number of distractors. We also utilize the area under the receiver operating characteristic (ROC) curve [11] to quantify target detection performance.

We conduct 10-fold cross validation for each subject. We train the SVM classifier, choosing the optimal kernel size  $k$  and cost parameter  $C$  for the SVM (from disjoint sets) that give the best validation performance. Validation performance is the average of the MFAR of nine classifiers, each of which is trained on a different nine-fold training set, and evaluated on a one-fold validation set.

## 4. GENERALIZATION PERFORMANCE

### 4.1 Parameter Selection

We conducted the 10-fold cross-validation to select the optimal parameters (kernel size  $k$  and cost parameter  $C$ ) for each subject. We exhaustively evaluated a variety of kernel size  $k$  and cost parameter  $C$  combinations during the validation. Figure 6 shows an example of the validation results for selecting the optimal parameters for subject #4. It shows the optimal kernel size to be  $k=10$  and  $C=1$ . We conducted the same procedures and obtained the optimal parameters for each subject.

### 4.2 Test Results

Having selected the optimal regularization parameters, we sought to estimate the detection performance on an *independent* test set not used for training or adjusting

regularization. We reported the averaged SVM results over ten runs as the final ERP-detection result to avoid using solutions from poor local optima.

We adopted the MFAR to evaluate the test performance. There are 22 sessions having the MFAR 0 to 2%. We observed that some targets did not receive button-click responses from subjects; which were mostly fake targets. We noticed that one particular target was consistently missed by all subjects (no button click following this target). We conjecture that some property (to be investigated) of this target makes it challenging to detect visually in the RSVP modality. After we removed the targets without button clicks, the averaged ranking results across 33 sessions improved from 12.17% to 8.09%. Figure 7 shows the histogram of the test result, we can see that 88% of the sessions (29 out of 33) have very low (less than 10%) MFAR and only four sessions have high false alarm rates. All the session achieve high ROC area ( $>0.9$ ) except these four sessions. In these four sessions, only one target in each session has low discriminant value, so it increases the MFAR. These results indicate that (assuming that targets would be uniformly distributed in the natural ordering of chips) approximately a 10-fold speed-up can be expected for the triaged-database compared to the original.

## 5. DISCUSSION

Our results show that the ERP-based image triage system provides a feasible solution for visual target search in image databases. The preprocessing procedure, such as disjoint windowing scheme provides effective training data and improves the detection accuracy. The expected performance in terms of high detection rate and low false alarms is limited by the imbalance between the prior probability of target and non-target images. Similarly, generalization is limited by the low samples-to-parameters ratio. Future work will focus on collecting more data from the general population in various image search contexts, and identifying robust discriminative low dimensional feature vectors for the ERP-based intent classification.

## ACKNOWLEDGEMENTS

This work was supported by DARPA and NGA under contract HM1582-05-C-0046 and by NSF under grants ECS-0524835, ECS-0622239, and IIS-0713690. It has been approved for public release, distribution unlimited. The data used in the experiments were collected at the Honeywell Human-Centered Systems Laboratory (Minneapolis, MN).

## REFERENCES

[1] S. Makeig, M. Westerfield, T P. Jung, S. Enghoff, J. Townsend, "Dynamic brain sources of visual evoked responses," *Science*, vol. 295, pp. 690–693, 2002.

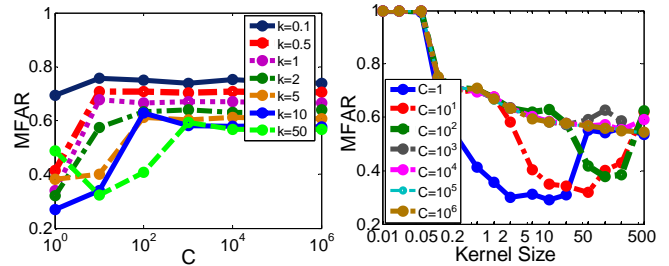


Fig. 6: The 10-fold cross validation result for parameter regularization on subject #4.

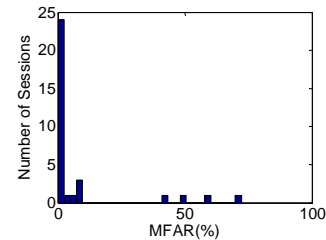


Fig. 7: Histogram of the test image sorting result on 33 test sessions from all subjects.

[2] M.D. Rugg, M.G.H. Coles, *Electrophysiology of mind: event-related brain potentials and cognition*, Oxford University press, Oxford, 1995.

[3] S. Thorpe, D. Fize, C. Marlot, "Speed of processing in the human visual system," *Nature*, vol. 381, pp.520-522, 1996.

[4] S. Mathan, S. Whitlow, D.Erdogmus, M. Pavel, P.Vervens, M. Dorneich, "Neurophysiologically driven image triage: a pilot study," in *Proceedings of the Conference on Human Factors in Computing Systems*, Montreal, Canada,1085-1090, 2006.

[5] A.D. Gerson, L.C. Parra, P. Sajda, "Cortically-coupled computer vision for rapid image search," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no.2, 174-179, 2006.

[6] Y. Huang, D. Erdogmus, S. Mathan, M. Pavel, "Comparison of linear and nonlinear approaches in single trial ERP detection in rapid serial visual presentation tasks," In *IEEE International Joint Conference on Neural Networks*, Vancouver, Canada, 2006.

[7] Y. Huang, D. Erdogmus, S. Mathan, M. Pavel, "Boosting linear logistic regression for single trial ERP detection in rapid serial visual presentation tasks," In *28<sup>th</sup> International Conference of IEEE EMBS*, New York, 2006.

[8] Y. Huang, D. Erdogmus, S. Mathan, M. Pavel, "A fusion approach for image triage using single trial ERP detection," In *3rd International IEEE Engineering in Medicine and Biology Society Conference on Neural Engineering*, Kohala Coast, HI, 2007.

[9] L.C. Parra, C. Alvino, A. Tang, B. Pearlmutter, N. Yeung, A. Osman, P. Sajda, "Single trial detection in EEG and MEG: keeping it linear," *Neurocomputing*, vol.52-54, pp.177-183, 2003.

[10] C.J.C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol 2, pp.121-167, 1998.

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