A Fusion Approach for Image Triage using Single Trial ERP Detection

Yonghong Huang, Student Member, IEEE, Deniz Erdogmus, Member, IEEE
Santosh Mathan, Member, IEEE, Misha Pavel, Member, IEEE

Abstract—**This paper addresses the problem of conducting visual target search on a large set of images. We present an approach that fuses neurophysiologic signals and overt physical responses to achieve high target detection accuracy. An experimental evaluation of the method was conducted using trained human experts in the paradigm of finding target objects in broad area aerial images. Spatial target likelihood maps were utilized to present estimated target locations in the images. Efficacy of the method was demonstrated on multiple subjects.**

I. INTRODUCTION

Information overload and lack of skilled human experts to process all acquired and stored data is an issue that emerged with the transition of the society to the information age. One such domain is image archive search for relevant/target-containing images in a large database of imagery, a situation that occurs in the fields, such as medical informatics and surveillance. In such domains, the existing imagery to be searched through exceeds the capacity of human experts using traditional search techniques. Therefore, effective image search through large volumes of images has become a crucial problem. One solution to this problem is an effective triage mechanism which would rapidly process the images and identify a subset that deserves careful inspection by a human expert. In complex domains, fully computerized machine learning solutions to the triage problem are not possible, although in more restricted scenarios such automated techniques have experienced success. Consequently, in this work, we exploited human experts as accurate target detectors and utilize their neurophysiologic and physical responses as indicators to see if an image is a target or a distractor.

The primary objective of our research is to develop an effective triage platform to boost the efficiency of image search. The key to achieve these dramatic boosts in efficiency lies in leveraging the information contained in electroencephalogram (EEG) signals with behavioral data. Recently, Thorpe and colleagues demonstrated that an evoked response potential (ERP), a brief change in the brain's electrical potential in response to critical events in an environment, can be a cue to detect targets with rapid serial visual presentation (RSVP) of images [1]. ERP can be estimated by analyzing EEG signals at a window size of hundreds of milliseconds (ms) following the stimulus onset. As shown in Figure 1 (by averaging 20 trials), ERP(P300) appears as a pronounced amplitude perturbation following the critical event in the EEG signals. The main challenge for ERP detection is the low signal to noise ratio. The conventional strategy for ERP detection is trail averaging, which compromises the efficiency of the image search and thus is infeasible for triage platform. Parra and colleagues recently developed promising approaches for single trial ERP detection [2,3], providing a solution to real-time brain computer interface design.

In our pilot study, we have investigated efficient machine learning algorithms to develop an ERP-based triage system using EEG signals [4-6]. The basic idea of an ERP-based triage system is to collect the EEG signals by monitoring the brain activities when a subject performed visual target detection from a huge amount of images and then detect ERPs associated to target responses. Our pilot study demonstrated that the ERP-based triage system is feasible and efficient for identifying targets within image sequences presented at RSVP modality. The experimental evaluation was carried out with subjects from the general population.

Based on our pilot study, we proposed a fusion solution for image triage, combining discriminative information from single trial ERP detection and overt physical response – a button press action by a subject following a target. The main benefits for the fusion method are based on the fact that the EEG and button responses provide good discrimination performance and thus we can take advantage of the strength of the two classifiers to reduce uncertainty. Classifier fusion has been widely used in a large variety of practical applications, such as remote sensing and biometrics personal identification [7,8]. The idea is to combine existing well performing classifiers to reduce overall classification error. Essentially there are two groups of classifier fusion techniques based on classifiers or classifier outputs. The methods operating on classifier outputs can be further divided into three types: class labels, class ranking and soft outputs. Among these three types, the soft output methods can be expected to produce the greatest improvement in classification performance [7]. Gerson and colleagues applied multiple linear classifiers and estimated optimal weights for each classifier using logistic regression, one of the class ranking fusion method for ERP detection [9]. They found out that the integration of neurophysiological and

Yonghong Huang is with the Computer Science and Electrical Engineering Department, Oregon Health and Science University, Portland, OR 97239 USA, (e-mail: huang@csee.ogi.edu).

Deniz Erdogmus is with the Computer Science and Electrical Engineering Department, Oregon Health and Science University, Portland, OR 97239 USA, (e-mail: derdogmus@ieee.org).

Santosh Mathan is with the Human Centered Systems Group, Honeywell Laboratories, Minneapolis, MN 55418 USA,
(e-mail: Santosh.Mathan@honeywell.com).

Misha Pavel is with the Biomedical Engineering Department, Oregon Health and Science University, Portland, OR 97239 USA,
(e-mail: pavel@bme.ogi.edu).

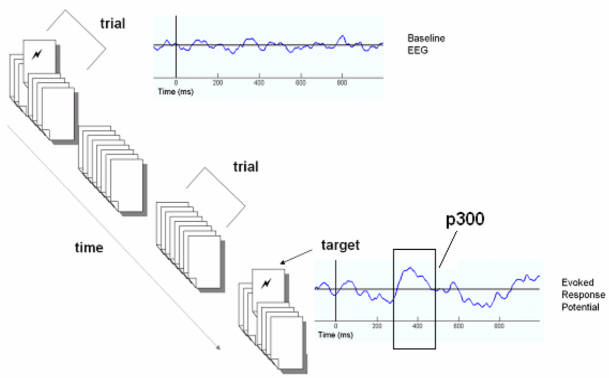


Figure 1. An illustration of the RSVP image presentation modality

behavioral responses can offer the best strategy for rapid image search.

In this work, we applied Bayesian fusion method, one of the soft output methods, to combine the ERP and button responses in conjunction with RSVP modality. The experiments were conducted on human professionals on target detection within large image sets. The objective was to investigate the efficacy of the fusion approach on skilled human experts.

II. METHODS

A. Data Collection

1) Data Acquisition

The data were collected from three professional image analysts: at the time of experiments, subjects S1, S2, and S3 had 6 years, 6 months, and 10 years of experience. None of them had experience with the RSVP modality. Subjects were instructed to perform target detection on broad satellite imagery, where the targets were predefined man-made structures.

As the subjects were trained in both the target characteristics and the RSVP modality prior to the tests, EEG and button data was collected to train the discrimination algorithms. The test broad area image was segmented into small chips at the same zoom level and resolution as the training images. All images were presented using the RSVP paradigm as illustrated in Figure 1. Images were presented in rapid succession for durations of 60 or 100 ms/image. Participants were asked to indicate the presence of targets by pressing a key as soon as they saw a target.

When the subjects conducted visual target identification, their EEG data was collected using the BioSemi system by connecting 32 electrodes on their scalps in the International 10-20 configuration. Timing triggers for all relevant events were recorded concurrently with EEG signals. The data were sampled at 256 Hz.

2) Data Preprocessing

There was no explicit bandpass filtering and the EEG data was segmented into *epochs*. Each epoch consisted of a short portion of EEG around each image onset (25 samples prior to and 255 samples following onset). Targets in the training set were limited to epochs where a button response was recorded within a second of the target trigger. The number of distractor samples was selected to match the number of target samples in the training session to avoid biasing classifier. In addition, computational and memory limitations prevented utilizing a large number of training samples due to the high data dimensionality. In the test sessions, *fake targets* that were not part of the original broad area image were introduced randomly in order to keep the subject alert and prevent boredom-related ERP-degeneration. Each channel in each epoch was rescaled to the interval [0,1] in order to emphasize location of ERP peak rather than its amplitude. The training and testing samples were limited to points between 195ms and 900 ms after the trigger. The data associated with each epoch were stored in a matrix (number of channels versus the number of samples of each epoch). The features used for classification were simply the temporal EEG measurements from 32 channels.

B. Classification Method

The fusion method is a combination of two classifiers: ERP-based classifier and button-based target likelihood estimate. Button press can provide accurate triage performance with higher latency and variability associated with motor response while ERPs can provide more precise localization of targets within RSVP context. The motivation of using the fusion approach is to combine the redundant source to reduce the false alarms.

1) ERP detection

Support vector machine (SVM) is a learning technique that has been widely used in pattern classification, especially in applications which involve large data sets with high dimensionality due to the implicit complexity penalization properties [10]. The basic idea is to find a class discriminating hyperplane in the feature space corresponding to the kernel with the largest margin. Support vectors are the data points lying at the boundary between the classes. Using the kernel trick, nonlinear decision surfaces can be obtained by solving a linearly constrained quadratic programming problem. We used a support vector machine (SVM) with radial Gaussian kernels as the ERP classifier in this work.

2) Button press detection

The button press latency following targets can be modeled by a gamma distribution whose parameters are estimated from the training session data using the maximum likelihood criterion. The obtained gamma distribution can then be utilized in the test session in a Bayesian manner to assess the probability that an image preceding a button press is the perceived target responsible for that action. These probabilities are obtained by integrating the gamma distribution model in the interval corresponding to each particular image preceding the button activity.

3) Fusion of Classifiers

Due to lack of extensive training data and the nonsmooth distribution of posterior probability estimates of ERP-SVM and button-gamma target likelihood evaluators, in this work,

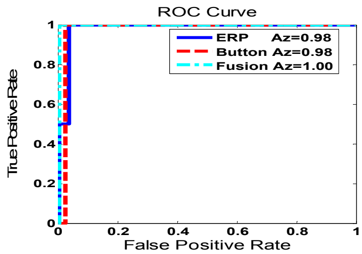
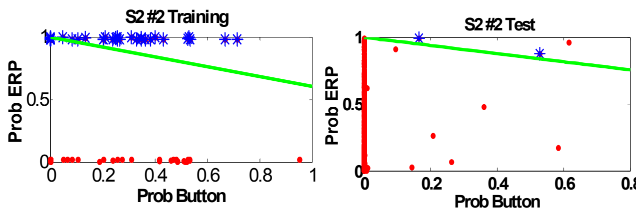


Figure 2. ROC curves for ERP, button press, and fusion



(a) (b)

Figure 3. Scatter plots ($P(target|\mathrm{ERP})$ vs $P(target|\mathrm{Button})$) for the training session (a) and test session (b). Green line is the boundary of the fusion classifier. Blue dots are target samples and red dots are distractor samples.

we have employed a linear classifier to accomplish the fusion of these two classifiers.

Ideally, one employs Bayesian fusion. The basic idea is to treat the posterior class-likelihood estimates of the various classifiers to be fused as components of a feature vector and building a Bayesian risk minimization discriminant for this particular vector. The output of such a Bayesian fusion model forms the optimal back-end-fusion strategy, yet it also faces all the challenges associated with designing truly Bayesian classifiers. Specifically in the binary case where there are two classes labeled {0,1}, if one uses multiple classification schemes to estimate $p_j(1|\mathbf{x})$, the posterior probability that a feature vector $\mathbf{x}$ belongs to class 1, or some function of this quantity, for $j=1,\ldots,J$ different classifiers, then one can construct a feature vector $\mathbf{z}$ where

$$\mathbf{z} = [p_1(1 \mid \mathbf{x}), \ldots, p_J(1 \mid \mathbf{x})]^T \quad (1)$$

and build a Bayesian fusion by constructing a Bayesian classifier to determine $p(1|\mathbf{x})=\int p(1|\mathbf{z})p(\mathbf{z}|\mathbf{x})d\mathbf{z}$.

Specifically, if one assumes that $p(1|\mathbf{z})\approx\boldsymbol{\alpha}^T\mathbf{z}+\beta$ around our operating point for the current $\mathbf{z}$ after ERP and button assessments are generated, we see that the Baysian fusion rule simplifies to

$$p(1|\mathbf{x})\approx \boldsymbol{\alpha}^T \int \mathbf{z}p(\mathbf{z}|\mathbf{x})d\mathbf{z}+\beta=\boldsymbol{\alpha}^T E[\mathbf{z}|\mathbf{x}]+\beta \approx\boldsymbol{\alpha}^T\mathbf{z}+\beta, \quad (2)$$

a simple linear combination rule whose weights can be optimized based on training data to minimize the classification error. Specifically, the fusion probability is obtained using the following:

$$P(target)=c_1P(target|ERP)+c_2P(target\,|Button)+c_3 \quad (3)$$

where $P(target|ERP)$ and $P(target|Button)$ are the two feature vectors. $P(target|ERP)$ is the ERP-classifier based probability that the image under consideration contains a target, and $P(target|Button)$ is the same probability as assessed by the gamma distribution model for response latency. The coefficients c_i $(i=1,2,3)$ are estimated by training a linear classifier to the probability estimates of the classifiers on the training data.

C. Evaluation

We used the area under the ROC curve to quantify the detection performance in terms of the relationship between false positives and true positives. The final performance was assessed using the area under the ROC curve.

III. Results

The primary focus of the experiments was to experimentally evaluate the efficiency of target search for experts using a combination of neurophysiological signals and overt physical responses in the context of the RSVP modality. Data were collected from three human experts (S1, S2, and S3). Each subject completed one training session with presentation rates of 100ms/image for s1 and s2 and 60ms/image for s3. In the test sessions, the three subjects completed one, four, and seven sessions respectively (Each test session is a different broad area image search). The duration of each image for S1 and S2 was 100ms. For S3, four sessions were at 60ms rate and three were at 100ms rate. The data was unbalanced at least a few hundred-fold, since there existed many more distractor images than target images. For S1 and S2, approximately 50 target images randomly positioned in a total of approximately 2750 images were presented in the training sessions. For S3, the training session contained 58 targets among 5492 distractors. All subject test sessions each contained two to eight targets within thousands of distractors. To avoid biasing the classifier, we randomly selected the same number of distractors as targets for training.

For ERP detection, the kernel width and slack parameters of the SVM were adjusted using cross validation and the SVM output was passed through a logistic function with unit slope at zero to normalize the output to [0,1] to facilitate their interpretation as probability levels. (Any monotonic function could be used and theoretically would not influence the final decision.) For button detection, we removed the fake target related button responses and obtained the gamma-distribution based target probability assignments for each test image clip. The fusion target likelihood assessment was obtained using the linear fusion rule in equation (3). Figure 2 shows the ROC curves of three detection schemes for S2 test # 1. The area under the curve for ERP detection, button and fusion are 0.98, 0.98 and 1. It is clear that the fusion approach achieved the best performance.

For fusion detection, we estimated a weighted linear classifier from equation (3) using training data and then employed on the test images as shown in Figure 3. Based on the estimated probability from the fusion classifier, a contour plot of the target likelihood was calculated for each image and overlaid on the actual broad area image for visual presentation to the human expert for final confirmation. Figure 4 illustrates a hotspot target likelihood assessment after fusion for a test image using S2 # 2. The four squares indicate the actual target locations for this particular image. This visualization technique allows efficient post processing of triage outputs.

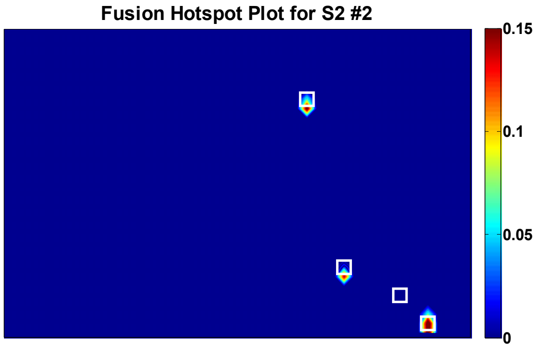


Figure 4. Contour plots of the probability distribution for s2 #2 test image containing four actual targets (white rectangles) and red indicates high likelihood assessment for that area containing a target (hotspot)

Table 1. ROC areas for all test sessions

		ERP	Button	Fusion
S1	Session 1	0.95	0.97	0.99
S2	Session 1	0.98	0.98	1
S2	Session 2	0.87	0.98	1
S2	Session 3	0.99	0.99	1
S2	Session 4	0.90	0.99	1
S3	Session 1	0.88	0.98	1
S3	Session 2	0.86	0.97	1
S3	Session 3	0.87	0.98	1
S3	Session 4	0.83	0.94	0.99
S3	Session 5	0.97	0.98	0.99
S3	Session 6	0.76	0.94	0.99
S3	Session7	0.94	0.99	1

Table 2. TP and FP for three subjects using fusion method

		# Targets	# of TP	# of FP
S1	Session 1	1	1	2
S2	Session 1	1	1	0
S2	Session 2	4	3	0
S2	Session 3	1	1	0
S2	Session 4	1	1	0
S3	Session 1	2	1	0
S3	Session 2	1	1	2
S3	Session 3	1	1	1
S3	Session 4	2	2	0
S3	Session 5	1	1	0
S3	Session 6	4	3	0
S3	Session7	2	2	1

Our experiments on three subjects have obtained high ROC area values as shown in Table 1. The true positive (TP) and false positive (FP) detection performances are shown in Table 2. The overall TP rate across subjects is 85.7% and the overall FP rate across subjects is 28.6%. The fusion process has been successful particularly in removing false positives, leading to an increase in performance. Note that the individual ERP and button classifiers already had very high area-under-curve measures, leading to the seemingly incremental increase in performance.

IV. Conclusion

We have implemented a neurophysiologic and behavioral response fusion technique to rapidly search large image databases for images containing targets of interest using the human visual system as the primary target detection sensor. The final prototype combining information from an ERP detection classifier and a behavioral response latency model was tested on data collected from three human subjects who are experts in searching for particular types of targets in the large image database domain that has been utilized in the experiments. Decision fusion from the two modalities has been observed to reduce false alarms significantly with respect to the ERP classifier and also with respect to the motor response alone. Overall, time savings of up to 87% (close to 10-fold increase in speed) for 100% detection level and 82% (close to 5-fold increase in speed) for 80% detection level compared with the baseline of manually searching for targets through the images using map survey software. Future work will focus on collecting more data from the general population in other image search contexts and identifying robust discriminative low dimensional feature vectors for ERP-based intent classification for general purpose brain computer interfaces.

Acknowledgment

This work was supported by DARPA under contract HM1582-05-C-0046 and by NSF under grants ECS-0524835 and ECS-0622239. It has been approved for Public Release, Distribution Unlimited. Some of the data used in the experiments were collected at the Honeywell Human-Centered Systems Laboratory (Minneapolis, MN).

References

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

[2] 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.

[3] A.D. Gerson, L.C. Parra, P. Sajda, “Cortical origins of response time variability during rapid discrimination of visual objects”, *NeuroImage,* vol. 28, no. 2, pp. 326-341, 2005.

[4] 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,” *IJCNN’06*, 2006.

[5] Y. Huang, D. Erdogmus, S. Mathan, M. Pavel, “Boosting Linear Logistic Regression for Single Trial ERP Detection in Rapid Serial Visual Presentation Tasks,” *EMBC’06*, 2006.

[6] S. Mathan, P. Ververs, M. Dorneich, S. Whitlow, J. Carciofini, D. Erdogmus, M. Pavel, Y. Huang, T. Lan and A. Adami, “Neurotechnology for Image Analysis : Searching for Needles in Haystacks Efficiently,” *AUGCOG’06*, 2006.

[7] D. Ruta and B. Gabrys, “An Overview of Classifier Fusion Methods,” *Computing and Information System*, vol. 7, pp. 1-10, 2000.

[8] L.I. Kuncheva, “ A Theoretical Study on Six Classifier Fusion Strategies,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, 2002.

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

[10] T. T. Frieb, N. Cristianini and C. Campbell, “The kernel-adatron algorithm: a fast and simple learning procedure for support vector machines,” *The 15th International Conference of Machine Learning*, 1998.