Neurocomputing 74 (2011) 2041–2051

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A framework for rapid visual image search using single-trial brain evoked responses

Yonghong Huang [a],*, Deniz Erdogmus [a,b], Misha Pavel [a], Santosh Mathan [c], Kenneth E. Hild II[a]

[a] *Biomedical Engineering Department, Oregon Health & Science University, Portland, OR 97239, USA*
[b] *Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA*
[c] *Human Centered Systems Group, Honeywell Laboratories, Redmond, WA 98052, USA*

ARTICLE INFO

Article history:
Received 7 December 2009
Received in revised form
1 November 2010
Accepted 9 December 2010
Communicated by S. Choi
Available online 21 March 2011

Keywords:
Brain computer interface (BCI)
Single-trial event-related potential (ERP)
Electroencephalography (EEG)
Image retrieval
Incremental learning

ABSTRACT

We report the design and performance of a brain computer interface for single-trial detection of viewed images based on human dynamic brain response signatures in 32-channel electroencephalography (EEG) acquired during a rapid serial visual presentation. The system explores the feasibility of speeding up image analysis by tapping into split-second perceptual judgments of humans. We present an incremental learning system with less memory storage and computational cost for single-trial event-related potential (ERP) detection, which is trained using cross-session data. We demonstrate the efficacy of the method on the task of target image detection. We apply linear and nonlinear support vector machines (SVMs) and a linear logistic classifier (LLC) for single-trial ERP detection using data collected from image analysts and naive subjects. For our data the detection performance of the nonlinear SVM is better than the linear SVM and the LLC. We also show that our ERP-based target detection system is five-fold faster than the traditional image viewing paradigm.

1. Introduction

Brain computer interface (BCI) provides a non-muscular avenue for the user to communicate with others and to control external devices. In the past decade, there has been a tremendous amount of research performed in the highly multidisciplinary field of BCI [18]. The first convincing demonstration of a direct functional interface between a brain and a robotic arm was documented in 1999 [5]. BCI is primarily applied to restore motor control for severely disabled people, particularly those suffering from spinal cord injury, amyotrophic lateral sclerosis, stroke, or cerebral palsy. The goal of BCI is to decode a user's intents, using only brain signals, in order to control an external device. There are a variety of methods used to record brain signals that might be used in an BCI. For example, electroencephalography, electrocorticogram, magnetoencephalography, functional magnetic resonance imaging, positron emission tomography. In reality, however, electrical field recording is more practical at present and in the near future [34]. The electrophysiological recording methods include: (1) electroencephalogram (EEG), which is recorded by electrodes on the scalp, (2) electrocorticogram, which is recorded by electrodes on the cortical surface, and (3) action potentials (from a single neuron or local field potentials), which are recorded by inserting electrodes into the cortex. Notice that EEG, unlike the last two invasive options, avoids the risks of brain surgery and is one of the most popular BCI approaches.

One way to use EEG for BCI involves extracting neural signatures from the data. Neural signatures, which are called event-related potentials (ERPs), are associated with perceptual and cognitive events. ERPs have drawn a lot of attention in the field of cognitive neuroscience [21,1,25]. A successful ERP-based BCI system depends on robust ERP detection, which can be very challenging due to the low signal-to-noise ratio of an ERP (the amplitude of a typical ERP is on the order of $1-10\ \mu V$, whereas the background EEG amplitude is on the order of $100\ \mu V$). The conventional strategy is to average across trials with identical stimuli, which increases signal-to-noise ratio and makes the ERP more detectable [14,20,32]. However, the trial-averaged approach filters out much of the information about cortical dynamics and requires that each stimulus be presented multiple times, which may not feasible for real-time systems. In order to avoid this limitation, single-trial methods have been recently developed. Instead of averaging across trials, one solution for single-trial approaches is to integrate information over sensors.

Object detection is one of the many possible applications for ERP-based BCI. A number of investigators have gained valuable insights into the mechanisms of object recognition and limits of

* Corresponding author.
E-mail addresses: huang@csee.ogi.edu (Y. Huang), erdogmus@ece.neu.edu (D. Erdogmus), pavel@bme.ogi.edu (M. Pavel), Santosh.Mathan@honeywell.com (S. Mathan), k.hild@ieee.org (K.E. Hild II).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2010.12.025

visual temporal processing [14,31,15]. Sajda et al. [27] have recently demonstrated an object (referred to as *target*) detection system, named cortically coupled computer vision, which uses single-trial ERP detection. They proposed to detect a subject's ERPs, which are elicited by rare, attended targets, while the subject views a stream of images presented at a high rate. This presentation paradigm is known as rapid serial visual presentation (RSVP). They then use a weighted linear logistic classifier (LLC) [23] for ERP detection. As an alternative to a brain evoked response, RSVP could be combined with a behavioral response, such as a button press, for a non-BCI approach. Moreover, a button press could be fused with a ERP to bring performance benefits [10,12]. However, comparing with a button press, the ERP-based BCI approach can be used by people with motor disorders, reduces fatigue, provides a continuous (as opposed to binary) measure of confidence, has a lower latency, and has lower variance in the timing. Therefore in this work we use ERPs as the only source. Fig. 1 shows example ERP signals associated with targets and distractors for one subject at channel Cz. The bottom traces are the EEG signals averaged across trials (trial averaging is shown here for visualization purposes only). When targets are presented, one can observe a perturbation in the EEG signal with a peak at 300 ms. Furthermore, there is no such discernable pattern when distractors are presented.

The main challenges of single-trial ERP detection are the high dimensionality and the scarcity of labeled EEG data. Although the existing methods [27,19,14,2] have been successful, they have several shortcomings. In particular, the algorithm for each observer must be trained anew for each session, and the system does not benefit from adding other observers. Most of the existing methods for single-trial ERP detection are trained using within-session data. The problem with using within-session data is that we may not record enough ERPs from a single subject in one sitting to sufficiently train the classifiers, at least in part because the amplitude of the ERP reduces for closely spaced targets. The natural tradeoff is that cross-session ERPs are expected to have considerably higher variation than within-session ERPs. Recently researchers start to explore cross-session training [16] and cross-subject training [28].

Here we develop an ERP-based BCI system for visual target search using RSVP based on incremental learning and cross-session training. We propose to use incremental learning as an alternative to batch learning [13] for ERP detection. The impetus for using incremental learning is to combine additional available training examples without having to retrain classifiers from scratch to reduce the computational load and memory storage, which is critical for real-time implementations. We also describe an adaptive training method that uses cross-session data. Fig. 2 shows the framework of our ERP-based BCI system. The experimental paradigm relies on the generation of ERP(s) in the frontal cortex of the subject who is instructed to look for specific kinds of scenes or target objects. The ERP is the brain response associated with the detection of a pattern matching a predefined target. When the subjects search for the objects in the image sequences in RSVP mode, their brain responds to the presentation of a relevant target scene/object and the EEG signals can be used to detect this cognitive process. The upper portion of the figure is the classification scheme. The stages include data collection, data extraction, ERP detection and image triage based on the ERPs associated with targets or distractors. The lower portion of the figure is the training schemes, which include the support vector machine (SVM) naive training using single-session data, the SVM batch learning using cross-session data and the SVM incremental learning using only support vectors.

This paper is organized as follows. First, we show that non-linear SVM has a better single-trial ERP detection performance than the linear SVM and the LLC on our data. Second, we demonstrate that our ERP-based target search system using within-session training has a throughput that is five times higher, in terms of square meters per unit time, than the traditional image viewing approach currently used by image analysts. Third, we demonstrate that the cross-session training approach is feasible on single-trial ERP detection problem, which could have large inter-session variances. Fourth, we introduce a novel incremental approach with less storage space and computational cost. We show that even though the incremental learning is as effective as batch learning, the memory storage is only 1/3 that of the batch learning (measured in terms of number of training samples) and the computational complexity is liner growth compared to the exponential growth of the batch learning.

2. Empirical data collection

The subjects performed target detection by clicking on a button (button presses in our experimental protocol were used to confirm targets explicitly recognized by subjects for proper data labeling) as soon as they saw a target. At the same time, we recorded their EEG signals. We used two computers to acquire data, one for image display and one for data collection.

2.1. RSVP image display paradigm

A large-scale satellite image (27 000 $\times$ 6500 pixels, representing an area of over 200 km^2) was decomposed into hundreds of smaller chips, which were labeled according to whether or not they contained a target. Each chip (500 $\times$ 500 pixels) represented an area of 0.09 km^2. The targets were surface-to-air missile sites found in the gray scale satellite imagery, as shown in Fig. 3. While very low level features like local textures are similar between

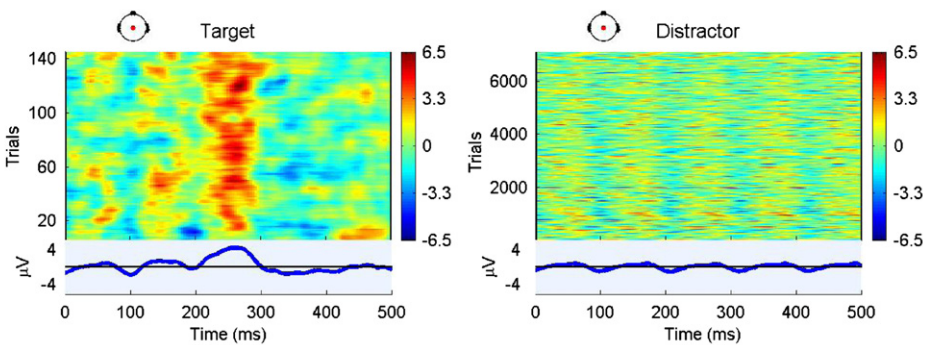


Fig. 1. Images of signals associated with targets (left) and distractors (right) respectively for one subject at channel Cz. The stimulus onset in each trial corresponds to 0 ms. The bottom traces are the EEG signals averaged over trials.

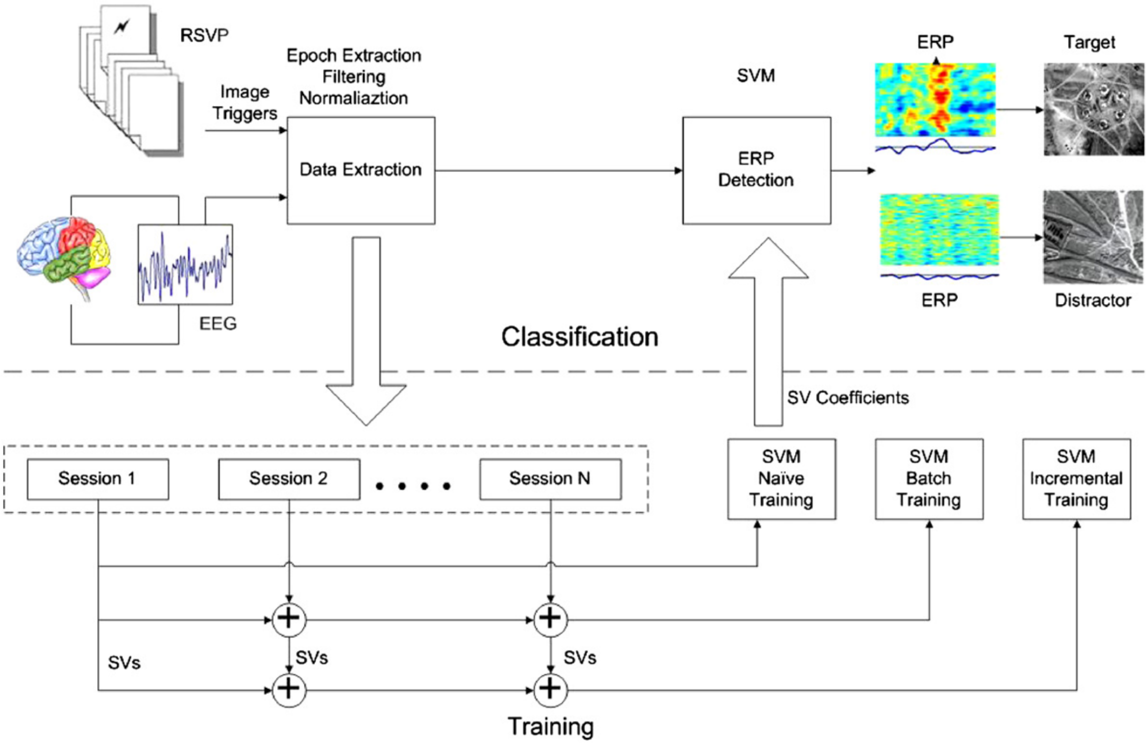


Fig. 2. The framework of the ERP-based image search system. The upper half portion is the classification scheme. The stages include data collection, data extraction, ERP detection and image triage. The lower half portion is the training schemes, which include the SVM naive training using single-session data, the SVM batch learning using cross-session data and the SVM incremental learning using only support vectors.

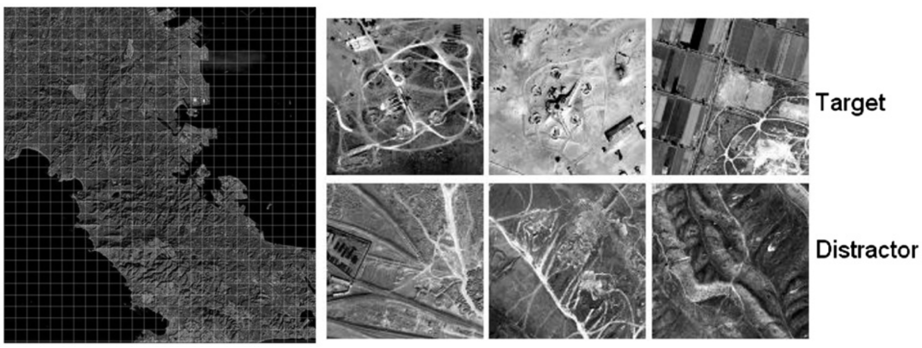


Fig. 3. A sample broad-area aerial image is segmented into hundreds of image chips (left). These chips are rapidly displayed to the subjects one-at-a-time. Target chips are encountered rarely (typically $\sim 1\%$). Examples of target and distractor chips are shown on the right.

target and nontarget image samples, target objects have larger scale correlations (e.g. specific layout of roads) – though non-systematically consistent – that can help a human expert/subject detect them accurately. In this study, the size of the targets was small and the scale, orientation, and position of the targets naturally varied. The subjects were shown multiple examples of missile sites and then asked to find other (similar) missile sites.

The image chips were presented using the RSVP paradigm. RSVP is an image presentation method previously used in behavioral studies [26,6]. During each RSVP session a sequence of images is displayed at a high presentation rate, such as 100 ms/image, as shown in Fig. 4. Recognition of a rare target image, which is embedded in a sequence of distracter images, elicits an ERP approximately 300 ms after the target stimulus onset. Images were presented in short bursts of approximately 5 s duration. To break monotony and minimize possible eye strain, consecutive blocks were separated by a fixation screen of user-controlled duration. The chips were presented on a 21 in CRT monitor using Presentation software (Neurobehavioral Systems, Albany, CA). Each image subtended 22 $\times$ 22 degrees of visual angle.

The RSVP sessions were structured in two phases, the training phase and the test phase. The stimuli were presented at the rate such as 100 ms/image controlled by the presentation experimental control software and the observers were instructed to respond as soon as they detected a target described above. Each subject was trained prior to testing. In the training phase, images were drawn with replacement from the image chip set and shown in a random order. Subjects received feedback on their responses at the end of each block. In the test phase, the chips were presented in the spatial order in which they occured in the broad-area image. There was no feedback in the test phase.

2.2. EEG acquisition

In the RSVP condition, EEG data were collected using a 32-channel BioSemi ActiveTwo system (BioSemi, Amsterdam,

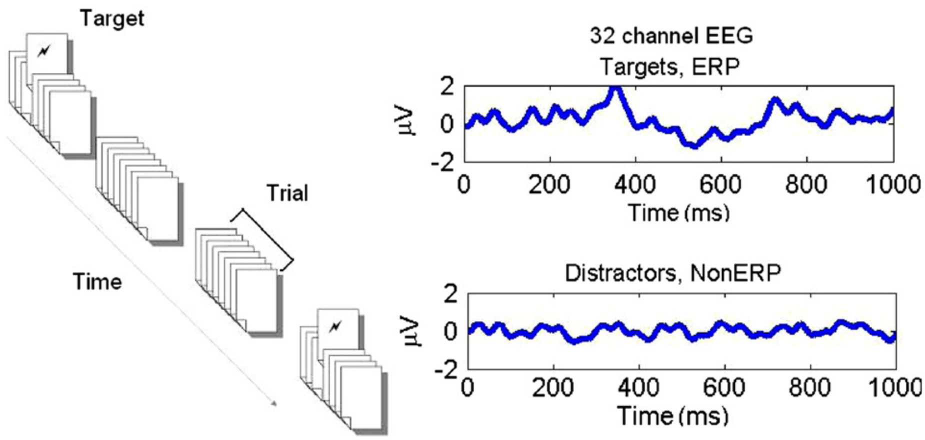


Fig. 4. An illustration of the RSVP image display modality (left). On the right, the upper trace is the trial-averaged ERP associated with the target stimuli (in one representative channel) and the lower trace is the trial-averaged ERP associated with the distracter stimuli. The stimulus onset corresponds to 0 ms.

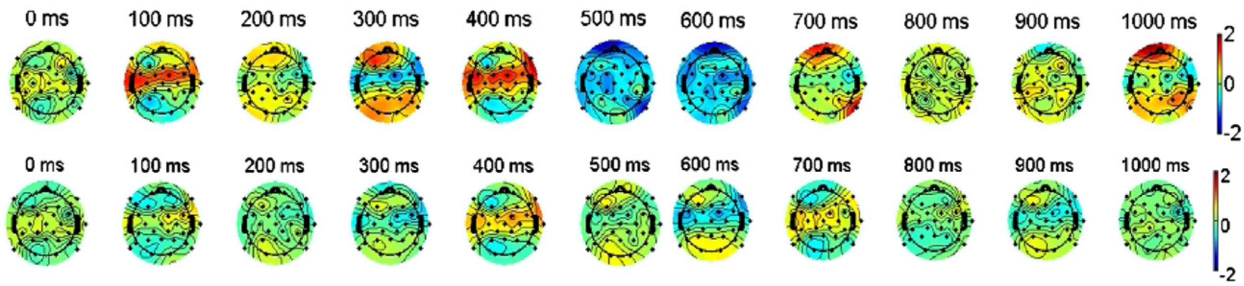


Fig. 5. Averaged ERP scalp distributions for a subject at 100 ms intervals following target (top) and distracter (bottom) stimulus onset. Notice that the spatiotemporal activity changes for the target trials and does not change for the distractor trials.

Netherlands). The data were sampled at 256 Hz. The presentation of each image was associated with a trigger generated 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 the EEG signals. The user's button presses, indicating the presence of perceived targets, were also recorded by the Biosemi system.

2.3. Three datasets

We used three datasets in this study. For the first dataset (IA Dataset #1), the subjects were intelligence analysts (IAs). Three professional IAs were recruited for the experiment. None of them had previous experience with the RSVP modality. The RSVP rate was 100 ms/image for subject 1 and 3 and 150 ms/image for subject 2. Each subject was trained on one (training) session and tested on seven (test) sessions.

For the second dataset (naive dataset), we used naive subjects. Four graduate students were recruited for the experiment. None of them had previous experience with the RSVP modality. To assess cross-session performance, we collected test data at different times and under different experimental conditions. Namely, data were collected from each subject during one morning session and one afternoon session each day on five consecutive days. Each session contained 200 trials. Each trial contained 37 images and was 5 s in length. Seventy five percent of the trials contained a single target instance. Images were drawn with replacement and shown in a random order. Each subject participated in 10 sessions in total.

For the third dataset (IA Dataset #2), the subjects were once again image analysts. Three professional IAs were recruited for this experiment, which compares the neurophysiologically driven target image search to the conventional broad-area manual image search. All subjects had experience with a broad range of imagery and target types. They were all trained in the use of geo-spatial analysis tools. None of them were familiar with the RSVP paradigm. The RSVP rate was 100 ms/image for subject 1 and 2. Subject 1 was tested on one session and subject 2 was tested on four sessions. Subject 3 was tested on seven sessions. The RSVP rates for this subject were 60 ms/image for training and four of the test sessions and 100 ms/image for the remaining three test sessions.

2.4. Data preprocessing

The raw EEG data were segmented into task-relevant epochs. Each epoch corresponded to a single image and consisted of a short segment of EEG (from the stimulus onset to 500 ms after the stimulus onset). Each epoch was associated with an image chip. The EEG was bandpass filtered between 1 and 45 Hz, using a 6th-order Butterworth filter to correct for DC drift and to limit the effects of 60 Hz electrical line noise. Based on other studies the peak latency of the recognition-related ERPs varies from 250 to 600 ms post-stimuli depending on stimulus and subject parameters [21]. Fig. 5 shows the electrical activity over the scalp as a function of time. One clearly discernible feature is a peak in the trial-averaged activity around 300 ms when target stimuli are present, whereas no amplitude change occurs for distractor stimuli. There are some amplitude changes around 700 ms, which are believed to be mainly caused by motor responses (button click activities). Based on our another study on the same set of stimuli, the averaged button response time (RT) across subjects is in the range of 500–600 ms with RT variance 20–50 ms depending on the stimuli difficulty for the RSVP rate of 100 ms/image.

To extract the neurally relevant portion of the ERPs (and to avoid EEG signals associated with motor responses, which are presumed to be unavailable in many practical applications), we utilized only the EEG signals from the stimulus onset to 500 ms post-stimulus. To normalize the EEG signals, we used the data from 100 ms before the stimulus onset to the stimulus onset. The data were normalized to have zero-mean and unit variance in each channel.

We used three steps to preprocess the training data. First, we removed targets that do not have an associated button click within 1.5 s of the stimulus onset (many studies in the literature have reported that the reaction time of a button press ranges from 300 to 500 ms depending on the complexity of the visual discrimination task [29]). We assumed that if there was no button press then there was no ERP. Second, we omitted all distractor samples within 1 s before or after the target stimuli so that the distractor windows were not contaminated by ERP leakage (hence, the windows are not contiguous when they are near target stimuli). Third, we created non-overlapping (*disjoint*) windows of EEG activity corresponding to distracters. The *disjoint* windowing scheme reduces the temporal correlation between data from two different windows. Each window was 600 ms in length and extended from 100 ms before the trigger (normalized window) to 500 ms after the trigger (epoch window). In non-overlapping windowing scheme, we discarded distractor samples occurring within the 600 ms window.

For test data, on the other hand, we used overlapping windows because we did not know the class label in advance. The 32-channel data in each 500 ms epoch were eventually concatenated to form a feature vector (32×129 dimension) and then the processed EEG measurements were subjected to the classifiers.

3. ERP-based target detection

Our goal in classification is to build a reliable ERP detector to accurately detect the brain responses associated with target stimuli. The inputs of the classifiers are raw EEG measurements and the outputs of the classifier are the likelihood values, which are used to label the EEG epochs according to whether or not they contain an ERP pattern.

3.1. ERP detector #1: linear logistic classifier

For the baseline ERP detector we adopt a linear discrimination approach based on a logistic regression model [11]. This method is based on the assumption that the EEG signals are a linear combination of distributed source activity and zero-mean white Gaussian measurement noise. The goal is to determine the linear projections of the EEG sensor measurements that maximize discriminability. To optimize the projection we assume that the probability of belonging to a class is represented by a logistic function. In particular the likelihood of sample belonging to the class follows a logistic model. To obtain the optimal weights, we use the batch gradient-descent algorithm and the least-square criterion. Given the linear projection and the corresponding class, we use the minimum overall Bayes risk as the criterion to determine the optimal threshold. Once we obtain the weights and the threshold from the training data, we can apply the linear detector to the test data to conduct the ERP detection. Note that the logistic classifier can also be trained using the maximum likelihood.

3.2. ERP detector #2: support vector machine

To explore more flexible classification strategy for high-dimensional data, we apply SVM on ERP detection. The SVM is a widely used statistical learning algorithm, especially for large data sets with high dimensionality [33,8,7,4]. Many researchers have reported that the SVM outperforms competing methods for their data [22,4,3].

The SVM algorithm is to map input observations to a high-dimensional feature space via kernel tricks and then optimize the decision boundary by constructing a maximum-margin hyper-plane. A hyper-plane is an affine subspace which divides the high-dimensional feature space into two half spaces, each of which is associated with one of the two classes. The optimization is a convex quadratic programming problem. After training, the optimal Lagrange multipliers and weights for each sample are obtained. The SVs, which are the data points lying at the border of the margin, have non-zero optimal solutions for their coefficients in the final discriminant, whereas the coefficients for the other data points converge to zero. Thus the training leads to a sparse non-parametric forward discriminant function. Once we train the SVM, we simply determine on which side of the decision boundary a given test ERP pattern lies and assign the corresponding class label to it. The parameters, such as kernel width σ^2, in the radial basis function (RBF) kernel, and the cost parameter C, which is a tradeoff parameter determining the relative weight of the penalty compared to the fit of the data, can be chosen by the users.

We applied the linear SVM and the RBF SVM on single-trial ERP detection. We adopted 10-fold cross-validation on the training session to adjust two regularization parameters of the RBF SVM: the kernel width of Gaussian kernels, σ^2, and the cost parameter, C, for each subject. We let the kernel size, σ^2, range from 0.01 to 500 and we let the cost parameter C range from 1 to 10^6. We vary σ^2 and C over the logarithmic grid formed by the selected values above. The SVM classifier is trained using the σ^2 and C giving the best validation performance.

3.3. Cross-session ERP detection

Given the low prior probability of a target in this task, we obtain very few ERP samples associated with targets within a long sequence of ERP samples associated with distractors because the amplitude of the ERP reduces for closely spaced targets. Since it is impractical to train a subject extensively in one session to obtain the sufficient ERP training samples, we train the classifiers using data aggregated across multiple sessions. Instead of conventional single-session training, we explore the feasibility of cross-session training on ERP detection to improve the generalization performance of classifiers. To assess cross-session performance, we use the RBF SVM as the cross-session ERP detector on the data collected at different times and under different experimental conditions.

3.3.1. Learning method #1: naive learning

We use the term "naive learning" to refer to single-session training. To compute the averaged single-session performance, we use only the current session as the test set and the previous session as the training set. For instance, we train on *session* $1(S_1)$ and test on S_2; then we trained on S_2 and tested on S_3; and so on until we trained on S_9 and tested on S_{10}. We compute the mean and the standard deviation across trials for the single-session performance.

3.3.2. Learning method #2: batch learning

We use the term "batch learning" to refer to cross-session training using all data from multiple sessions. In batch learning, we aggregate the data across all previous sessions for training. For instance, We train on $S_1 \cup S_2$ and test on S_3; and so on until we train on $S_1 \cup \cdots \cup S_9$ and test on S_{10}. The aggregated data are subjected to the classifier to evaluate the cross-session performance. We use the RBF SVM with kernel size $k = 1$ and cost parameter $C = 10$ as the ERP detector to evaluate the cross-session performance.

We use Monte Carlo method to repeat pseudorandom sampling to compute the averaged cross-session performance. To simulate a realistic scenario, we use only the current session as the test set and the previous sessions as the training set to create Monte Carlo pseudorandom sessions. In batch learning using cross-session data, we train on $S_1 \cup S_2$ and test on S_3; train $S_2 \cup S_3$ and test on S_4; and so on until we train on $S_8 \cup S_9$ and test on S_{10}. We compute the mean and the standard deviation across trials for the double-session performance. Similarly we use the Monte Carlo method to create pseudorandom sessions for multiple-session training and computer the averaged performance.

3.3.3. Learning method #3: incremental learning

The cross-session training in batch mode may produce higher performance than the single-session training due to more training samples. However, such batch training is computationally intensive and it is infeasible for real-time systems. Incremental learning paradigm, as opposed to the batch learning paradigm in which all training examples are provided at once for optimization, is a training mode where only a few training examples are added at a time to update model parameters. The motivation of incremental learning is to deal with very large training sets or non-stationary data. An important advantage of incremental learning is that it allows the algorithm to combine additional available training examples without having to retrain classifiers from scratch. This has numerous benefits, including saving a substantial amount of storage space and speeding the computation up. One of the main difficulties with using incremental learning methods is the sensitivity of choosing training parameters.

The essential property of the SVM algorithm is that only the SVs contribute to the decision boundary so that the remaining training examples may be regarded as redundant. Based on this property, Syed et al. proposed an incremental learning for the SVM to deal with large datasets [30]. They segmented a huge dataset into small partitions to avoid problems associated with limited available memory, and incrementally trained the SVM with the small partitions. Their results demonstrated that the SVs selected by the SVM algorithm was a minimal set. Any further removal of data samples significantly deteriorated the performance because the loss of SVs led to loss of vital information about the class distribution.

Motivated by Syed's method, we developed an incremental learning scheme for cross-session ERP detection. The basic idea of the incremental learning ERP detection is to train an SVM on the previous session of EEG data. The SVs found during training are preserved and combined with the training samples from the current session. For the cross-session EEG data in Section 2.3, there were 10 datasets, S_1–S_{10}. Instead of training on all previous data as S_1, $S_1 \cup S_2, \ldots, S_1 \cup \cdots \cup S_9$ as in the batch learning, we preserve the SVs from the previous training sets and discarded the redundant data. Let V_i represent the SVs in session S_i. For the proposed incremental learning scheme, we train S_1 preserve V_1, and test on S_2; train S_2, preserve $V_1 \cup V_2$, and test on S_3; …; train S_8, preserve $V_1 \cup V_2 \ldots \cup V_8$ and test S_9 respectively.

3.4. Evaluation methods

We adopted the area under the receiver operating characteristic (ROC) curve (AUC) as the performance evaluation criterion for ERP detection. The ROC curve [24] describes the relationship between the false positive fraction (FPF) and the true positive fraction (TPF) as the threshold for discrimination between two classes is varied. We used Delong's non-parametric approach [9] to compare correlated AUCs by generating an estimated covariance matrix.

4. Comparison of ERP detection techniques: SVM vs. LLC

We conducted an experiment to select a classifier on single-trial ERP detection for incremental adaption process. The LLC was compared to a linear and a nonlinear SVMs. We evaluated the classification performance on both professional image analysts and naive subjects.

We compared three classifiers – the RBF SVM, linear SVM, and LLC – on both IA Dataset #1 and the naive dataset. Our null hypothesis, H_0, is that there is no statistically significant difference between the two correlated AUCs (the AUC of the SVM and the AUC of the LLC). The optimal parameters for the three IA subjects are $\sigma^2 = 0.05$ and $C = 100$; $\sigma^2 = 0.1$ and $C = 100$; $\sigma^2 = 10$ and $C = 10{,}000$, respectively. The optimal parameters for the four naive subjects are $\sigma^2 = 1$ and $C = 10$; $\sigma^2 = 1$ and $C = 10$; $\sigma^2 = 1$ and $C = 100$; $\sigma^2 = 1$ and $C = 1000$, respectively. We used $C = 1$ for the linear SVM (in our experience the performance of the linear SVM is not sensitive to the C value). Based on our 10-fold cross-validation results, the parameter C and σ^2 could be sensitive, especially when the value C is small. Therefore we suggested wide range of parameters. We also noticed that they were sensitive across subjects, so we performed cross-validation parameter search accordingly for each subject.

For the IA subjects the AUC of the RBF SVM is higher than the LLC by 0.107, 0.103 and 0.095, respectively, as shown in Fig. 6. The two-tailed p values for the three IA subjects are $<<0.0004$

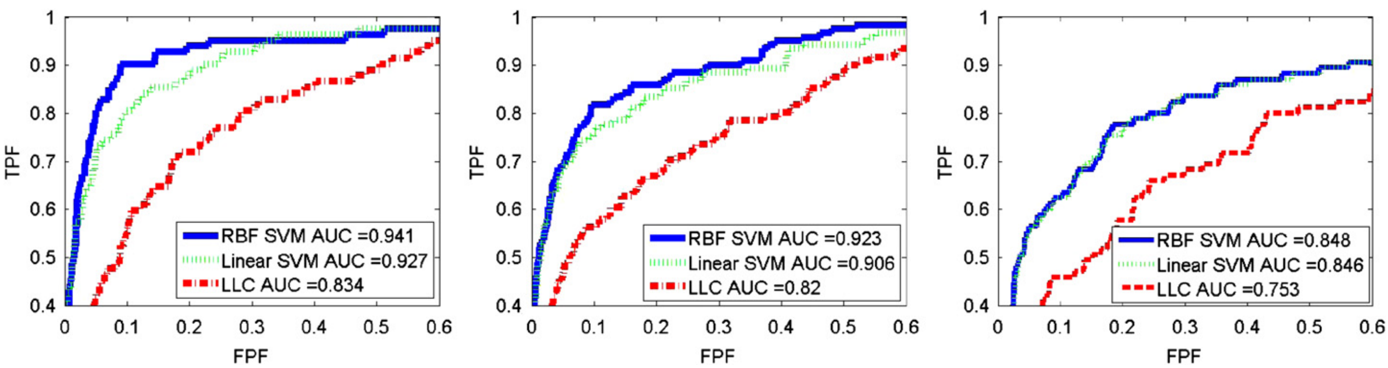


Fig. 6. The ROC curves of the RBF SVM (solid), linear SVM (dotted), and LLC (dash-dotted) for the three IA subjects from IA Dataset #1. ROC curve depicts the relationship between the false positive fraction (FPF) and the true positive fraction (TPF). The performance in term of the area under the ROC curves (AUC) for each classifier is shown.

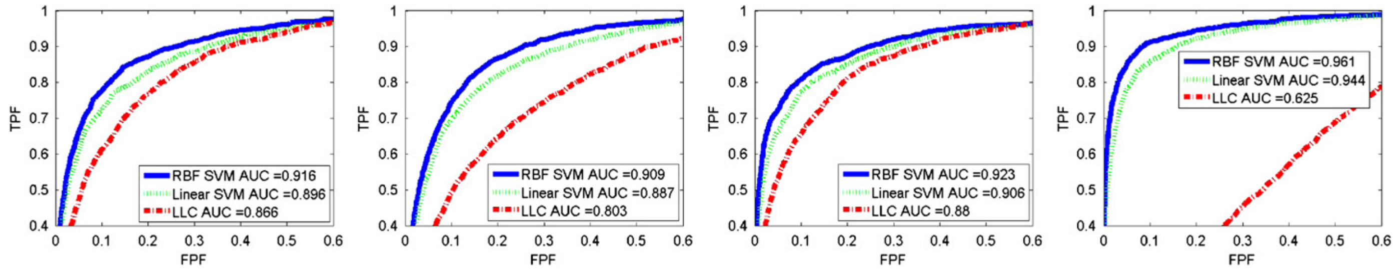


Fig. 7. The ROC curves of the RBF SVM (solid), linear SVM (dotted), and LLC (dash-dotted) for four naive subjects. ROC curve depicts the relationship between the false positive fraction (FPF) and the true positive fraction (TPF). The performance in term of the area under the ROC curves (AUC) for each classifier is shown.

(z = 5.145, 5.718 and 4.550, respectively). For the four naive subjects the AUC of the RBF SVM is higher than the LLC by 0.050, 0.106, 0.043 and 0.336, respectively, as shown in Fig. 7. The two-tailed p values for four naive subjects are $<<0.0004$ ($z = 14.529$, 21.063, 12.008 and 47.420, respectively). From the results of both the IA dataset and the naive dataset, one can conclude that the difference between the AUCs from the RBF SVM and the LLC is highly statistically significant and thus we can reject H_0.

Similarly, in the comparison of the linear SVM and the LLC, the two-tailed p values for the three IA subjects and the four naive subjects are $<<0.0004$ ($z = 5.259$, 4.773, 4.383, 8.708, 17.560, 7.413, and 14.271, respectively). One can conclude that the difference between the AUCs from the linear SVM and the LLC is also highly statistically significant and thus reject H_0.

Comparing the RBF SVM with the linear SVM, the two-tailed p values for the IA dataset are 0.230, 0.023, 0.021 ($z = 1.202$, 2.280, 2.308, respectively) and the two-tailed p values for the naive dataset are $<<0.0004$ ($z = 9.420$, 10.237, 9.848, and 44.137, respectively). The difference between the AUCs from the RBF SVM and the linear SVM is statistically significant for all subjects except one IA subject.

For our data, the non-parametric test for correlated AUC measurements indicates that the linear and nonlinear SVMs achieve significantly better performance, in a statistical sense, than the LLC for ERP detection. Likewise, the RBF SVM, at the cost of additional computational complexity, performs better than the linear SVM for ERP detection. The result indicates that the data may not be linearly separable in the original feature space so that the nonlinear projection to the high-dimensional feature space helps the discrimination of the two classes. However the tradeoff is the computation efficiency.

For high-dimensional EEG data (32×129 dimension[1]) and sparse training samples (for IA dataset, roughly 50 positive samples and 500 negative samples; for naive dataset, roughly 140 positive samples and 950 negative samples), the SVMs are capable to capture nonlinear class separation boundary. The SVMs map input data to a high-dimensional feature space via kernel tricks and optimize the linear separating hyper-plane in feature space. However, the LLC is simply a linear hyper-plane boundary. This could be the main reason for performance discrepancy. Some regularization could be made to the LLC to approximate the SVMs, for instance, by modified iteratively re-weighted least squares estimation procedure or a modified penalized log likelihood function. Consequently, based on these comparison results, we chose the RBF SVM as the classifier for the incremental adaptation process. This cross-model performance comparison should not be taken as universal conclusion.

[1] The previous work on feature dimension reduction and extraction using time–frequency and wavelet transforms did not increase classification performance on our data. In the later studies, we did dimensionality reduction especially in the channel direction, which described in a separate paper. In this paper our focus is incremental learning. Here we use the original data as features.

5. Efficiency: comparison of the ERP approach and the tradition approach

We conducted a second experiment to explore the efficiency of neurophysiologically driven image search. More specifically, we compared the ERP approach with the tradition image viewing approach on the detection rate (the ratio of the number of detected targets and the number of total targets) and detection speed (the ratio of the total search time and the total image area, which is measured in units of s/km^2). The study was carried on a group of professional image analysts.

The ERP-based target detection system collects EEG signals as a subject observes a high-speed scan of the thousands of chips extracted from several broad-area images and then analyzes the data to identify ERPs. We used IA Dataset #2 in this experiment. We employed the RBF SVM as the ERP detector in this study. Based on the estimated likelihood values from the SVM, we constructed a contour plot of the target likelihood and overlay it on the associated broad-area image. The human experts used this plot to do a final confirmation. In the tradition image viewing approach, participants used a geo-spatial analysis tool called GlobalMapper (Global Mapper Software). It provides zoom and pan controls and allows high resolution satellite imagery to be efficiently searched and annotated. Participants were allowed as much time as they wished to search the targets in a broad-area image. A set of prototype images depicting the targets were shown to each participant.

Fig. 8 shows the target likelihood contour maps (for all contours above a fixed threshold) for one of the broad-area images used in the test. One can see that, in this case, the ERP-based approach accurately detects one target and has a few false alarms. This visualization technique allows efficient post-processing of the triage outputs.

Fig. 9 shows the detection rate and the detection speed averaged across test sessions for each subject. We can see that the averaged detection rates of the ERP system are equal to or higher than those of the manual method. The detection speeds of the ERP system are much faster than those of the manual method for all subjects. The overall detection rate across subjects for the ERP system is 93% compared to 67% for the manual approach. The averaged AUC across seven test sessions for the three subjects is 0.82. The overall false alarm rate for the ERP system is higher than the manual method, which had zero false alarms. The precision for the ERP system is 78% and the precision for the manual method is 100%. The overall detection speed for the ERP system is 5.3 times faster than the manual method.

6. Cross-session ERP detection

We conducted a third experiment to extend the results from the first experiment and to explore the robustness of the ERP-detection approach to changes of the EEG signals over time on all

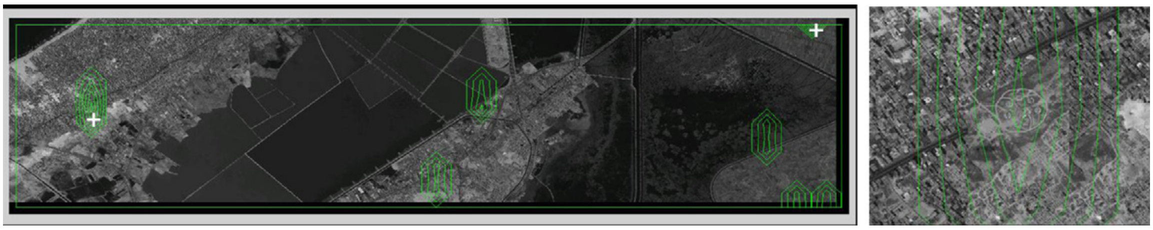


Fig. 8. Broad-area image with overlaid contour maps (left). The crosses indicate true target locations. Users can zoom into the contour hotspot to confirm the presence of a target (right).

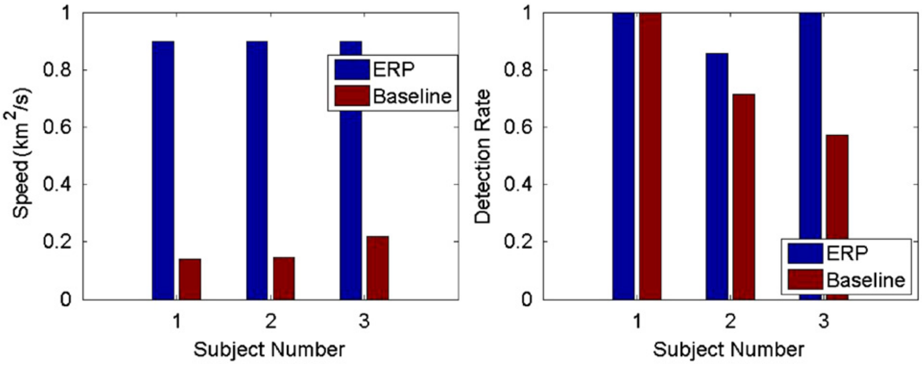


Fig. 9. The detection speed (left) and detection rate (right) averaged across test sessions for the three IA subjects.

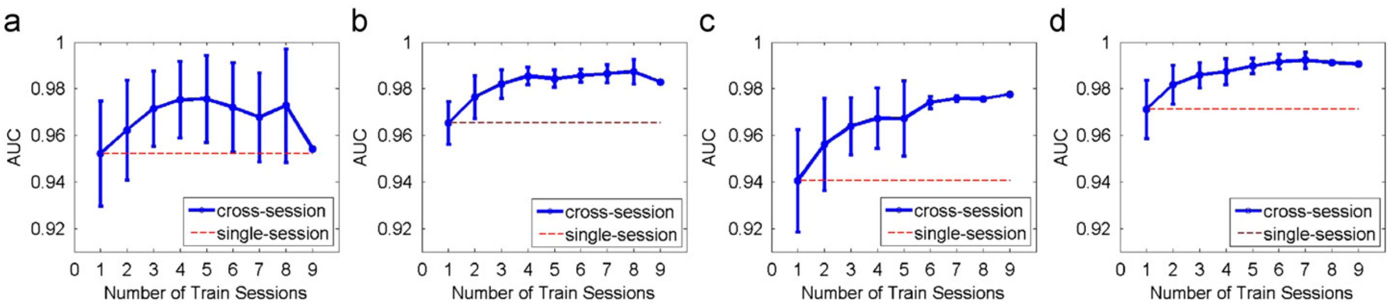


Fig. 10. The cross-session performance in term of the area under ROC (AUC) on different number of training sessions using the RBF SVM. The AUC as a function of the number of training sessions for four subjects in naive learning using single-session data (red dashed line) and batch learning using cross-session data (blue solid curve). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

four naive subjects. Here we evaluate (1) the viability of cross-session batch learning on ERP detection by comparing it with naive learning and (2) the efficiency of incremental learning on ERP detection by comparing it with batch learning. We evaluated the method on cross-day data from a naive subject group.

6.1. *Batch learning vs. naive learning*

Fig. 10 shows the ERP detection performance in term of AUC using the batch learning on cross-session data and the naive learning using single-session data. To simulate a realistic scenario, we use only the current session as the test set and the previous sessions as the training set to create Monte Carlo pseudorandom sessions. The averaged performance is computed across Monte Carlo trials. The GKSVM achieves high cross-session generalization performance for all subjects. The averaged AUC exhibits a generally increasing trend with the inclusion of additional training data from subsequent sessions for four subjects. The averaged performance across subjects increases around 5%. Our results demonstrate the viability of cross-session single-trial ERP detection. Given a reasonable amount of cross-session training data, the SVM achieves excellent generalization performance, as indicated by the high AUC values. The results suggest that the classifier is able to capture much of the range of variation in our EEG data and approaches the problem of characterizing between session variation in signal statistics. The results demonstrate that inter-session variances do not significantly deteriorate detection performance.

6.2. *Incremental learning vs. batch learning*

We compared the single-trial ERP detection performance on the batch learning method, and the incremental learning method. Based on our 10-fold cross-validation results on the RBF SVM, the parameters showed consistent performance across sessions for each subject in our data. Therefore we selected the same parameters across sessions in the batch learning and the incremental learning. Fig. 11 shows the cross-session ERP detection performance in terms of the AUCs. The batch learning SVM, which uses all previous cross-session data, achieves higher AUCs than the naive learning SVM for all sessions of three of the subjects. The incremental learning SVM achieves similar AUCs as the batch approach for all subjects. The AUCs of both incremental learning and batch learning exhibit a generally increasing trend with the inclusion of additional training data from subsequent sessions for all four subjects.

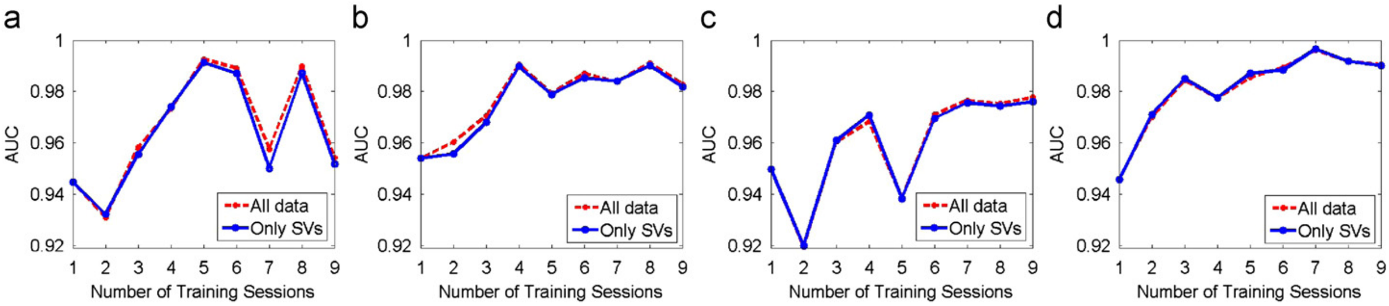


Fig. 11. The incremental learning performance in term of the area under ROC (AUC) using the RBF SVM. The AUC as a function of the number of training sessions for four subjects for batch learning (red dashed, using all previous data for training) and incremental learning (blue solid, only the SVs are propagated). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

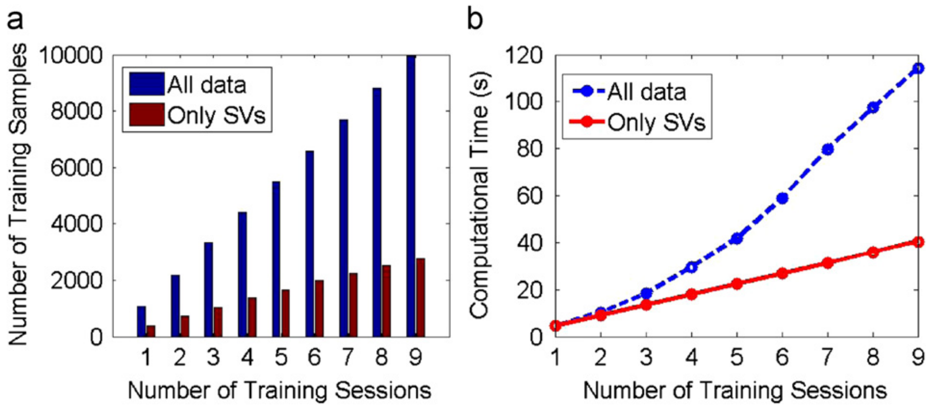


Fig. 12. The memory storage in terms of the number of training samples (a) and the computational time (b) for various number of training sessions available for subject 1 using batch learning (blue, uses all previous data for training) and incremental learning (red, only the SVs are propagated). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Next we compare the memory storage and the computational elapsed time of the methods. Fig. 12 shows the memory storage in terms of the number of training samples and the computational elapsed time for subject 1 for both incremental learning and batch learning. The other three subjects have similar results. As one can be seen in these plots the incremental learning costs much lower storage space and computational load. The memory storage of the incremental learning is only 1/3 that of the batch learning (measured in terms of number of training samples) and the computational complexity of the incremental learning is liner growth compared to the exponential growth of the batch learning.

Our results on the cross-session dataset for the previously described realistic scenario demonstrate the viability of cross-session single-trial ERP detection. Given a reasonable amount of cross-session training data, the SVM achieves excellent generalization performance, as indicated by the high AUC values. The results suggest that the classifier is able to capture much of the range of variation in our EEG data. The results demonstrate that inter-session variances do not significantly deteriorate detection performance. The incremental learning method performs as well as the batch mode due to the fact that only the SVs contribute to the decision boundary. Since the incremental learning compacts the previous training data to the SVs and then incorporates only the SVs with the new dataset, it is more computationally efficient than the batch learning method.

7. Conclusion

We extend the Sajda et al.'s work [27] and demonstrate the ERP-based target detection system for speeding up visual target search by tapping into the split-second perceptual judgments of humans. We propose cross-session training for single-trial ERP detection and demonstrate the efficacy of incremental learning on cross-session data. The incremental learning method using only SVs performs better than the naive method and it achieves, with a substantially lower memory storage and computational cost, a performance similar to the batch method for cross-session ERP detection. The linear and nonlinear SVM classifiers significantly outperforms the LLC for single-trial ERP detection due to good generalization capabilities of SVM. The ERP-based approach provides a higher detection rate and speed, albeit with more false alarms, than the tradition image viewing approach. It should be noted that many of the false positives can be removed with limited effort by manual confirmation of the prospective targets.

This work represents only a first step towards our vision of neurophysiologically driven image triage system, and much work remains. Next we will investigate the characteristics of ERP patterns and explore some statistical models, such as the mixed-effects models [17].

Acknowledgments

This work was supported by DARPA under contract HM1582-05-C-0046 and by NSF under grants ECS-0524835, ECS-0622239, and IIS-0713690. Satellite images included in this paper were produced by DigitalGlobe Inc., Longmont, CO 80501, USA (c) 2003. The two IA datasets used in the experiments were collected at the Honeywell Human-Centered Systems Laboratory (Minneapolis, MN). The author would like to thank Tan Lan for help with the naive data collection. It has been approved for public release, distribution unlimited. The views, opinions, and/or findings contained in this article/presentation are those of the author/presenter and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Department of Defense.

References

[1] K.B. Campbell, E. Courchesne, T.W. Picton, K.C. Squires, Evoked potential correlations of human information processing, Biological Psychology 8 (1) (1979) 45–68.

[2] N. Bigdely-Shamlo, A. Vankov, R.R. Ramirez, S. Makeig, Brain activity-based image classification from rapid serial visual presentation, IEEE Transactions on Neural Systems and Rehabilitation Engineering 16 (5) (2008) 432–441.

[3] B. Scholkopf, A.J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, New Jersey, NY, 2001.

[4] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2 (1998) 121–167.

[5] J.K. Chapin, K.A. Moxon, R.S. Markowitz, M.A. Nicolelis, Real-time control of a robot arm using simultaneous recorded neurons in the motor cortex, Nature Neurosciences 2 (1999) 664–670.

[6] M.M. Chun, M.C. Potter, A two-stage model for multiple target detection in rapid serial visual presentation, Journal of Experimental Psychology: Human Perception and Performance 21 (1) (1995) 109–127.

[7] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (1995) 273–297.

[8] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge University Press, Cambridge, UK, 2000.

[9] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (1988) 837–845.

[10] A.D. Gerson, L.C. Parra, P. Sajda, Cortical origins of response time variability during rapid discrimination of visual objects, Neuroimage 28 (2) (2005) 342–353.

[11] D.W. Hosmer, S. Lemeshow, Applied Logistic Regression, Wiley-Interscience, 2000.

[12] Y. Huang, D. Erdogmus, S. Mathan, M. Pavel, A fusion approach for image triage using single trial ERP detection, in: Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering, Kohala Coast, HI, 2007, pp. 473–476.

[13] Y. Huang, D. Erdogmus, S. Mathan, M. Pavel, Detecting EEG evoked responses for target image search with mixed effect models, in: Proceedings of the 30th International IEEE EMBS Conference, Vancouver, BC, Canada, 2008, pp. 4988–4991.

[14] J.S. Johnson, B.A. Olshausen, Timecourse of neural signatures of object recognition, Journal of Vision 3 (2003) 499–512.

[15] C. Keysers, D. -K Xiao, P. Foldiak, D.I. Perrett, The speed of sight, Journal of Cognitive Neuroscience 13 (1) (2001) 90–101.

[16] M. Krauledat, M. Tangermann, B. Blankertz, K.R. Muller, Towards zero training for brain-computer interfacing, PLoS ONE 3 (8) (2008) e2967.

[17] N.M. Laird, J.H. Ware, Random-effect models for longitudinal data, Biometrics 38 (1982) 963–974.

[18] M.A. Lebedev, M.A.L. Nicolelis, Brain-machine interfaces: past, present and future, Trends in Neurosciences 29 (9) (2006) 536–546.

[19] S.J. Luck, An Introduction to the Event-Related Potential Technique, The MIT Press, Cambridge, MA, 2005.

[20] S. Makeig, M. Westerfield, T.P. Jung, S. Enghoff, J. Townsend, Dynamic brain sources of visual evoked responses, Science 295 (2002) 690–694.

[21] B.S. Oken, Evoked Potentials in Clinical Medicine, LippincottRaven, Philadelphia, PA, 1997 Chapter 15.

[22] E. Osuna, R. Freund, F. Girosi, Support vector machines: training and applications, Technical Report AIM-1602, 1997.

[23] L.C. Parra, C.D. Spence, A.D. Gerson, P. Sajda, Recipes for the linear analysis of eeg, Neuroimage 28 (2) (2005) 326–341.

[24] M.S. Pepe, The Statistical Evaluation of Medical Tests for Classification and Prediction, Oxford University Press, New York, NY, 2003.

[25] T.W. Picton, The p300 wave of the human event-related potential, Journal of Clinical Neurophysiology 9 (4) (1992) 456–479.

[26] M.C. Potter, E.I. Levy, Recognition memory for a rapid sequence of pictures, Journal of Experimental Psychology 81 (1) (1969) 10–15.

[27] P. Sajda, A.D. Gerson, M.G. Philiastides, L.C. Parra, Single-trial analysis of EEG during rapid visual discrimination: enabling cortically coupled computer vision, Towards Brain-Computer Interfacing, MIT Press, 2007.

[28] P. Sajda, L.C. Parra, C. Christoforou, B. Hanna, C. Bahlmann, J. Wang, E. Pohlmeyer, J. Dmochowski, S.-F. Chang, In a blink of an eye and a switch of a transistor: cortically-coupled computer vision, Proceedings of the IEEE 98 (3) (2010) 462–478.

[29] E. Schoroger, A. Widmann, Speeded responses to audiovisual signal changes result from bimodal integration, Psychophysiology 35 (1998) 755–759.

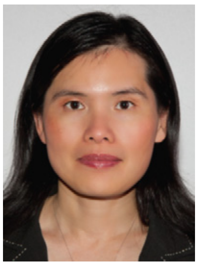
[30] N.A. Syed, H. Liu, K.K. Sung, Incremental learning with support vector machines, in: Proceedings of the International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 1999.

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

[32] R. VanRullen, S.J. Thorpe, The time course of visual processing: from early perception to decision-making, Journal of Cognitive Neuroscience 13 (4) (2001) 454–461.

[33] V.N. Vapnik, Statistical Learning Theory, John Wiley & Sons Inc., New York, NY, 1998.

[34] J.R. Wolpaw, Brain-machine interfaces for communication and control: a mini-review, Clinical Neurophysiology 57 (2004) 607–613.



Yonghong Huang received her Ph.D. degree in Biomedical Engineering from the Oregon Health & Science University in 2010. During her interdisciplinary Ph.D. program, she studied machine learning, signal processing, and computational neuroscience. She is currently employed at OHSU as a postdoctoral researcher. She has been performing research on brain evoked responses and brain-computer interfaces since 2005. Previously, she received her B.S. degree in Electrical Engineering from the South China University of Technology. In 2005, she received her M.S. in Electrical Engineering from the University of New Brunswick in Canada, where she performed research on myoelectric control of limb prostheses. She was awarded the three-year Postgraduate Federal Scholarships by Natural Sciences and Engineering Research Council of Canada for 2005–2008. Her research interests are machine learning and signal processing with the applications to computational biology for clinical medicine and health research. She has a handful of peer-reviewed publications and has been served as a reviewer for several journals and conferences.

Deniz Erdogmus graduated with B.S. in EE and the B.S. in Mathematics in 1997, and M.S. in EE in 1999 from the Middle East Technical University, Ankara, Turkey. He received his Ph.D. in ECE from the University of Florida in 2002, where he stayed as a postdoctoral research associate until 2004. He is currently an Assistant Professor in the ECE Department at Northeastern University. Prior to this position he held the same rank jointly at the CSEE and BME Departments of the Oregon Health and Science University. His research focuses on statistical signal processing and machine learning with applications to cognitive signal processing including brain-computer interfaces and technologies that collaboratively improve human performance. He has over 60 journal publications and he has been serving as an associate editor for IEEE and Elsevier journals in these areas.

Misha Pavel is a Professor and a head of Biomedical Engineering with a joint appointment in Medical Informatics and Biomedical Computer Science at Oregon Health & Science University. He is the Director of Point of Care Laboratory focused on unobtrusive monitoring, neurobehavioral assessment and computational modeling of behavioral and cognitive functions with applications to care for aging and chronically ill populations. Previously, Dr. Pavel was a Technology Leader at AT &T Laboratories, where he was developing networked, wireless and mobile applications for information access and context aware interactions. As a faculty at New York University and at Stanford University, Dr. Pavel worked on sensor fusion, modeling of pattern recognition in sensory-motor systems and human–computer communication systems. As a Member of Technical Staff at the Bell Laboratories, he was developing new approaches to network analysis and modeling by incorporating characteristics of human behavior. Dr. Pavel received his Ph.D. in Experimental/Mathematical Psychology from New York University; M.S. in Electrical Engineering from Stanford University, and B.S. in Electrical Engineering from Polytechnic Institute of Brooklyn.

Santosh Mathan is a researcher in the Human Centered Systems Group at Honeywell Laboratories in Redmond, Washington. His research focuses on the development of techniques for estimating cognitive state in challenging real-world application contexts. He is currently principal investigator on the DARPA funded Neurotechnology for Intelligence Analysts Program. He is also exploring the feasibility of using sensor-based workload estimates for usability evaluation. Santosh has a Ph.D. in human–computer interaction from Carnegie Mellon University.

Kenneth E. Hild II is a senior member of IEEE. He received a Ph.D. in Electrical Engineering from the University of Florida, Gainesville, where he studied feature extraction, information theoretic learning, and blind source separation in the Computational NeuroEngineering Laboratory. Later, he studied biomedical informatics at Stanford University, Palo Alto. From 1995 to 1999, he worked as an Advisory Development Engineer in the Advanced Concepts group at Seagate Technologies. From 2000 to 2003, he taught several graduate-level classes on adaptive filter theory and stochastic processes at the University of Florida. From 2003 to 2007, he was at the Biomagnetic Imaging Laboratory at UCSF, San Francisco, where he applied variational Bayesian techniques for biomedical signal processing of encephalographic and cardiographic data. He is currently employed at the Oregon Health & Science University, Portland, where he is performing research on brain computer interfaces and both human and computer vision.