

UC San Diego

UC San Diego Previously Published Works

Title

A Brain-Computer Interface (BCI) for the Detection of Mine-Like Objects in Sidescan Sonar Imagery

Permalink

<https://escholarship.org/uc/item/6js697qs>

Journal

IEEE Journal of Oceanic Engineering, 41(1)

ISSN

0364-9059

Authors

Barngrover, Christopher
Althoff, Alric
DeGuzman, Paul
et al.

Publication Date

2016

DOI

10.1109/joe.2015.2408471

Peer reviewed

A Brain–Computer Interface (BCI) for the Detection of Mine-Like Objects in Sidescan Sonar Imagery

Christopher Barngrover, Alric Althoff, Paul DeGuzman, and Ryan Kastner

Abstract—Detection of mine-like objects (MLOs) in sidescan sonar imagery is a problem that affects our military in terms of safety and cost. The current process involves large amounts of time for subject matter experts to analyze sonar images searching for MLOs. The automation of the detection process has been heavily researched over the years and some of these computer vision approaches have improved dramatically, providing substantial processing speed benefits. However, the human visual system has an unmatched ability to recognize objects of interest. This paper posits a brain–computer interface (BCI) approach, that combines the complementary benefits of computer vision and human vision. The first stage of the BCI, a Haar-like feature classifier, is cascaded in to the second stage, rapid serial visual presentation (RSVP) of images chips. The RSVP paradigm maximizes throughput while allowing an electroencephalography (EEG) interest classifier to determine the human subjects' recognition of objects. In an additional proposed BCI system we add a third stage that uses a trained support vector machine (SVM) based on the Haar-like features of stage one and the EEG interest scores of stage two. We characterize and show performance improvements for subsets of these BCI systems over the computer vision and human vision capabilities alone.

Index Terms—Boosting, brain–computer interface (BCI), mine-like object (MLO), object detection, rapid serial visual presentation (RSVP), sidescan sonar.

I. INTRODUCTION

THE detection and classification of mine-like objects (MLOs) in sidescan sonar imagery is a problem of grave importance to the safety of our military. The manual approach is still the primary technique for finding and eliminating these objects. Even with the aid of underwater robotics to capture data, the task of processing the imagery is very time consuming. The automation of these tasks would save time and money, but the dynamic underwater environment makes this a difficult task for classifiers, with the human operators still leading in performance.

Manuscript received May 28, 2014; revised October 07, 2014; accepted February 23, 2015.

Associate Editor: J. Cobb.

C. Barngrover with the Department of Computer Science, University of California San Diego, La Jolla, CA 92093 USA, and also with the Space and Naval Warfare (SPAWAR) Systems Center Pacific (e-mail: cbarngrover@ucsd.edu; chris.barngrover@navy.mil).

A. Althoff and R. Kastner are with the Department of Computer Science, University of California San Diego, La Jolla, CA 92093 USA (e-mail: aalthoff@ucsd.edu; kastner@ucsd.edu).

P. DeGuzman is with the Neuromatters, LLC, New York, NY 10038 USA (e-mail: pdeguzman@neuromatters.com).

Digital Object Identifier 10.1109/JOE.2015.2408471

There is an extended history of research on the topic of automated detection of mine-like objects in sonar imagery. Much of the earlier research uses a model based approach with knowledge of the target, focusing on highlights and shadows created by the strong reflection and occlusion of the protruding MLO from the seafloor [1]–[3]. In some research, the model for various range regions is used as a matched filter, convolving the filter with the image to detect regions of interest [4].

Some recent research has looked at ways to improve the training process based on manipulating the training data. One such approach uses the partially observable Markov decision process (POMDP) to learn what additional views will be beneficial to the training [5]. Another uses the knowledge of an imbalance between targets and similar clutter to improve the training with the infinitely imbalanced logistic regression (IILR) algorithm [6]. There is also an approach that preprocesses the data using the manifold learning technique called Diffusion Maps and analyzes the features that result [7].

The advancements in the capabilities of sonars and the autonomous underwater vehicles utilizing them has led to research using machine learning techniques and well known computer vision features. These features do not have a concept of the target model and instead focus on local descriptors within a window of the image. The machine learning algorithms require large training data sets to optimize a classifier.

One machine learning capability, called AdaBoost, was altered to be a feature selection process, choosing Haar-like features from a pool of feature options [8]. The research paper by Viola and Jones was proposed as a face detector, but it has been applied to many other targets [9]–[11]. This concept has also been applied to MLO detection in sonar, but using a variation of boosting called GentleBoost [12], [13].

While computer vision and machine learning approaches to MLO detection have made significant progress over the past decade, the human visual system's ability to recognize objects of interest remains unmatched. Humans can easily and robustly identify objects in a scene, regardless of the scale, lighting, background clutter, etc. Moreover, when an image is flashed quickly, humans are able to ascertain the *gist* of a scene in as little as a few hundred milliseconds [14]. However, when tasked to process large databases of images, computers have the advantage over humans in terms of processing speed, data throughput, and absence of fatigue.

Many variations of BCIs have previously been used to tackle the image search problem [15]–[19]. One particular BCI system, Cortically-Coupled Computer Vision (C3Vision), synergistically combines the respective advantages of computer vision and human vision. C3Vision starts with a computer

vision system that preprocesses large images into smaller *image chips* around potential regions of interest (ROI). This is done using a model-based approach with the assumption that targets are known *a priori*. This framework contains a feature dictionary, containing extracted low-level features, which is used to infer objects using a grammar-based reasoning engine [17]. C3Vision then presents these image chips using the rapid serial visual presentation (RSVP) paradigm, which maximizes throughput while still retaining the ability to decode neural activity related to detection and recognition as measured by electroencephalography (EEG). This RSVP component detects neural attention changes rather than behavioral responses, such as button presses.

For example, in one application the C3Vision system performed well on satellite imagery. In this example, intelligence analysts search for targets of interest, such as surface-to-air missile sites or air fields, within large images on the order of several hundred gigapixels in size. The analysts can rapidly view thousands of image chips and identify ROIs, which deserve closer inspection. The C3Vision BCI setup showed that the search process could be accelerated without degraded detection performance [16].

A. Our Approach

The C3Vision BCI setup has performed well in the past on various data sets using a simple computer vision method to prepare image chips. And for the application of detecting MLOs in sidescan sonar imagery, we have seen that the GentleBoost Haar-like feature selection algorithm works well [12]. We propose creating a BCI system that uses a trained Haar-like feature classifier to create the image chips and the RSVP capability of C3Vision for human processing.

Our approach begins by independently considering an implementation of the Haar-like feature selection algorithm to create a classifier and an experimental setup using the C3Vision RSVP capability with human subjects to rank images of interest. Then the primary experiment of this paper will be to cascade the Haar-like feature classifier with C3Vision's RSVP to show the improvements over the individual capabilities.

We also consider an additional BCI system, which has three stages. The first two stages are the same as the previous approach, starting with the same Haar-like feature classifier with output processed by subjects using C3Vision RSVP. The final stage is a support vector machine (SVM) classifier trained using a feature vector composed of the same selected Haar-like features as well as EEG *interest score* features.

There are four major contributions of this paper:

- the concept of RSVP with EEG systems and human subjects to the application of detecting MLOs in sidescan sonar imagery;
- the introduction of a new BCI system using a Haar-like feature classification stage cascaded in to an RSVP human classification stage for the detection of MLOs in sidescan sonar imagery;
- an SVM classifier whose training feature vector consists of both Haar-like features and EEG interest score features;
- a three-stage BCI system with the first two stages being computer vision to create image chips before RSVP-based

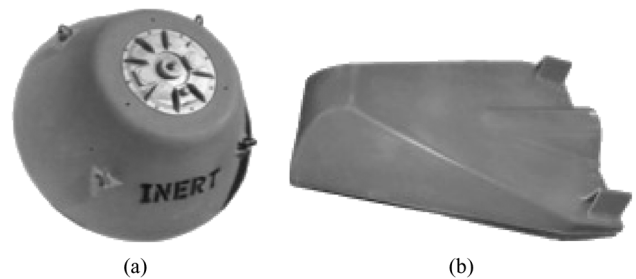


Fig. 1. Examples of the inert mine types contained in the data sets used for this experiment. These objects sit on the sea floor protruding up from the bottom. There are ten different Type 1 mines and seven different Type 2 mines included in our data. (a) Type 1., (b) Type 2.

human processing, followed by a third classification stage using an SVM.

The remainder of this paper is organized as follows. Section II introduces the sonar images data set and the MLO targets. Then we describe the Haar-like feature classifier and its performance on our data set in Section III. Next, in Section IV, we explain the rapid serial visual presentation (RSVP) paradigm used and its performance on our data set. Section V proposes a BCI setup that cascades the Haar-like feature classifier before RSVP, showing three experiments using variations of the computer vision classifier. A novel BCI step is presented in Section VI that uses the Haar-like features and the EEG interest scores to train a support vector machine (SVM) classifier, which is used as a third stage following the same two stage setup from Section V. Finally we conclude in Section VII.

II. DATA SET

The data used for this paper were collected using remote environmental monitoring Units (REMUS) vehicles in collaboration with the Space and Naval Warfare Systems Center Pacific (SSC-PAC) in San Diego, CA, USA. The REMUS vehicle is equipped with two Marine Sonic sidescan sonars operating at a frequency of 900 kHz. The missions were all ran at an altitude of four meters producing images with a 30 m range from each sonar. The combined sonar image used in this research is 1024 by 1000 pixels.

There are two different types of mines, shown in Fig. 1, placed in test fields in San Diego Bay. Multiple passes of the various mine locations, including ten truncated cones (Type 1) and seven stealth wedges (Type 2), produce the many looks included in our data. The primary data set consists of 450 sonar images containing 150 target mines. Additionally, the Haar-like feature classifier in the next section was previously trained on a separate data set of 975 images each containing an MLO, with 426 truncated cones and 549 stealth wedges. This training data set was collected in the same manner and in the same fields of San Diego Bay [13].

The data used in this research are relatively easy for this task, but are congruent with the environments currently considered in real world mine clearing situations. MLO detection in the more cluttered sea bottom environments is an important task, but we focus on relatively clutter free bottoms in the very shallow water (VSW) environments for this research. The data

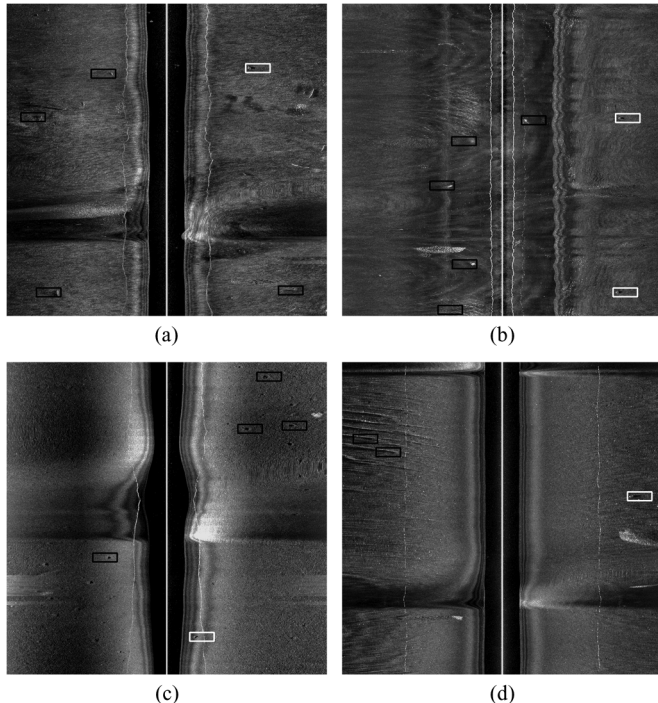


Fig. 2. Sonar images show the minor complexity of this data set. Images (a), (c), and (d) contain results from the vehicle making a 180 degree turn. Image (a) shows some rock clutter and other objects, image (b) shows a poor quality image with highlight clutter, and image (c) shows small holes or dimples in the sea floor. Image (d) shows a sand ripple bottom type. The images were processed by the Haar-like feature classifier introduced in Section III and show white boxes around the true positives (TPs) and black boxes around the false positives (FPs).

set does have some complexity, however, including sand ripples, image quality, rock clutter and the prominent vehicle turn regions as shown in Fig. 2. The images in the figure were processed by the Haar-like feature classifier presented in the next section and the white boxes show the correctly found MLO, while the black boxes show false positives (FPs).

This data set was collected via a star pattern over a target, causing most images to have at least one turn. These turns are evident in Fig. 2 images (a), (c), and (d). Image (a) shows rocks on the sea floor, while image (b) shows low quality with highlight clutter spots. Image (c) shows small holes or dimples in the sea floor, which look like shadows. Finally, image (d) shows some sand ripples on the sea floor. All of these slight complexities and clutters are common in this data set.

The Haar-like feature classifier in the next section and RSVP-based classifier in the following section both use all 450 images as a testing set. Section V also uses this same 450 image data set as a testing set for the the BCI system including a Haar-like feature classifier cascaded in to an RSVP stage. Finally, Section VI divides this data set in to 225 training images to create the new SVM classifier and 225 testing images to test the three stage BCI setup.

III. HAAR-LIKE FEATURE CLASSIFIER

The Haar-like feature classifier used in this research is produced using the feature selection algorithm proposed by Viola and Jones [8] and later applied to MLOs by Sawas *et al.* [12]. The classifier is composed of a series of stages, each of which

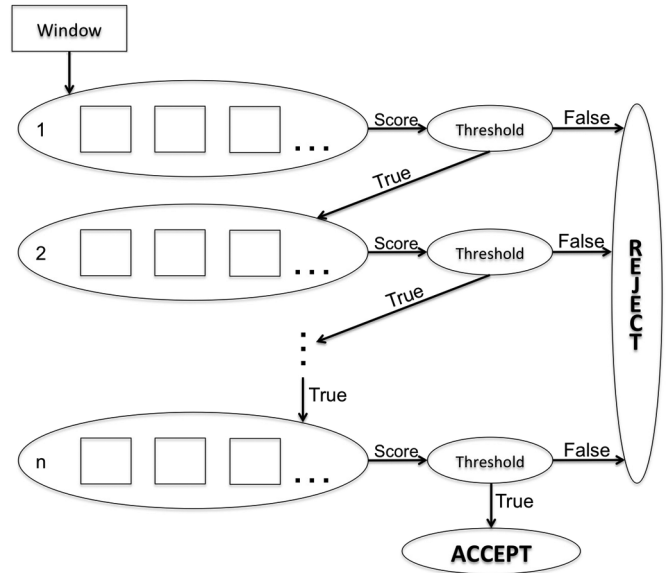


Fig. 3. Schematic view of a cascaded classifier, with the large ovals to the left being stages and the squares within them being features. Each classifier stage produces a score that is tested against a threshold to determine if the stage is passed. If any stage threshold is not passed then the window is rejected. If all stages are passed then the window is accepted.

is a classifier in itself, which are cascaded to speed up the processing time and improve accuracy. This is a sliding window classifier with a fixed window size and a horizontal step of one pixel and a vertical step of two pixels. Fig. 3 shows the cascade concept, with each stage composed of multiple features represented by squares, and a positive output from one stage required for entrance in to the next stage.

The algorithm is created using a boosted feature selection process to select the features of each stage and to determine when a stage is complete. The pool of features from which the algorithm selects includes variations of the Haar-like feature. Both the feature selection algorithm and the feature itself are described in more detail in the following two subsections. In Subsection III-C we present the experimental results of this classifier on our data set.

A. GentleBoost Feature Selection

The GentleBoost feature selection algorithm combines improvements in boosting algorithms with the Viola and Jones [8] proposal for feature selection to improve the classifier optimization process. In a boosting algorithm, each iteration estimates based on the current classification scheme and then updates the classifier based on the error from ground truth. The rule to update the classification scheme is based on the weighting of the training data. This update to the weights is what changes between versions of boosting algorithms. All versions use training data (x_i, y_i) where x_i is the data and $y_i = \{-1, 1\}$ is the label for positive or negative.

The GentleBoost variation of boosting uses adaptive Newton steps to update the classification rule rather than the more volatile log-ratio update of the previous Real AdaBoost method [20]. Compared to the potentially large updates in Real AdaBoost, this allows for smaller changes to weight distribution at a given update step, which dampens large shifts due to difficult

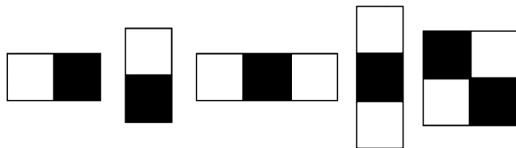


Fig. 4. Visualization of the five Haar-like features that we use in this paper. The sum of the pixels in the white rectangle is subtracted from the sum of the pixels in the black rectangle to calculate the Haar-like feature value.

training examples and leads to a more stable convergence. The classification output is a real value where the sign gives the classification and magnitude gives the confidence. The weight update function is shown in (1). Instead of taking the log-ratio of the probabilities given by the classification hypothesis, $h_t(x)$, the more stable difference is used

$$w_i \leftarrow w_i \exp [y_i \times (1 - 2h_t(x_i))]. \quad (1)$$

The feature selection component involves a change to the traditional boosting algorithm, where each iteration restricts its consideration to one feature at a time. The process operates by iteratively selecting new features until the goal true positive rate (TPR) and false positive rate (FPR) are met. The TPR is the number of correctly labeled targets out of the total targets in the training data, while the FPR is the number of incorrectly labeled targets out of the total number of negative windows considered. In the case of our training, the goal TPR is 99.5% and the goal FPR is 50%. These thresholds are on a per stage basis, so that each stage should maintain a high rate for true positives but reduce the false positives by at least half. The algorithm will continue adding stages to the cascade until a set number stages are created, fourteen in this research, or an acceptance ratio is achieved.

B. Haar-Like Feature

The pool of features from which the GentleBoost feature selection algorithm creates the classifier consists of variations of the Haar-like feature. This feature is based on the Haar wavelet and captures changes in pixel intensity between neighboring regions. There are many variations of the feature but this classifier only considers five basic types, which are visualized in Fig. 4.

The actual feature is calculated by taking the difference between the sum of pixel intensities in the white regions and the sum of the pixel intensities of the black regions shown in the visualizations. A particular Haar-like feature consists of the feature type, the location within the window being considered, and the size of the rectangles. The pool of features considered by the feature selection algorithm includes all possible variations of these attributes for our fixed window size of 79 by 29 pixels, which amounts to 2 54 ,145 Haar-like features.

C. Results

The Haar-like feature classifier we use in this paper was previously trained by the GentleBoost feature selection algorithm on another data set collected under the same exact parameters as the data presented in Section II. The training set used to train the classifier in previous research consists of 975 sonar images, each containing an example mine target, with 426 truncated cones

and 549 stealth wedges [13]. This classifier consists of seven stages and 36 total features. Here we present how this classifier performs on our data set of 450 sonar images with 150 mine-like targets.

The receiver operating characteristic (ROC) curve representing the performance of this classifier on the data set is presented in Fig. 5. The vertical axis shows the true positive rate (TPR), which is the number of targets correctly labeled out of the 150 possible targets in the data set. The horizontal axis shows the false positives per square kilometer (FP per km²), which is the number of times the algorithm incorrectly labeled a window as a target in terms of a square kilometer of the sea floor. We are able to produce this value based on the known range of the sonar images in this data set, which gives the area covered in a given image.

The points that make up the curve are created by processing the data set with the classifier under a varying threshold value. This threshold value is the number of positive window neighbors necessary to mark a window as positive. For example, if the threshold value is zero then any window that is classified as positive by the cascade is output as positive. Alternatively, if the threshold is two, then a window that is positively classified must have two other neighboring windows also classified as positive for the output to be positive. In other words, the higher the threshold, the more conservative the classifier is about outputting a positive label for a window. Fig. 5 highlights three points on the curve where the threshold is set to zero, two, and four, respectively.

IV. RAPID SERIAL VISUAL PRESENTATION

The human visual system has the ability to comprehend the *gist* of scenes when presented with images in rapid succession at a rate of five to ten images per second. Using this rapid serial visual presentation (RSVP) paradigm allows for the maximization of image throughput while maintaining the ability to decode electroencephalography (EEG) signals related to moments of visual interest.

The C3Vision system, used in this research, is a brain-computer interface (BCI) that uses an RSVP setup designed to distinguish between two brain states: moments of visual interest triggered by positive image chips containing something salient versus idle moments when negative image chips do not contain anything of particular interest. It is important to note that the system does not decode brain signals based on what exactly the user sees in an image, which would be a very difficult problem to solve, but rather marks when in time a subject detects something of interest. Also, the RSVP system detects neural attention changes rather than behavioral responses, such as button presses.

To discriminate between positive and negative examples, an EEG interest classifier must be calibrated for an individual subject, since each subject has a different EEG marker for a moment of visual interest. Therefore, a short calibration session is required for each user of the RSVP setup in C3Vision. During the calibration, sets of known targets and nontargets are presented to the user in RSVP to calibrate the EEG interest classifier using the hierarchical discriminant component analysis (HDCA) algorithm. This algorithm linearly combines EEG electrodes in such

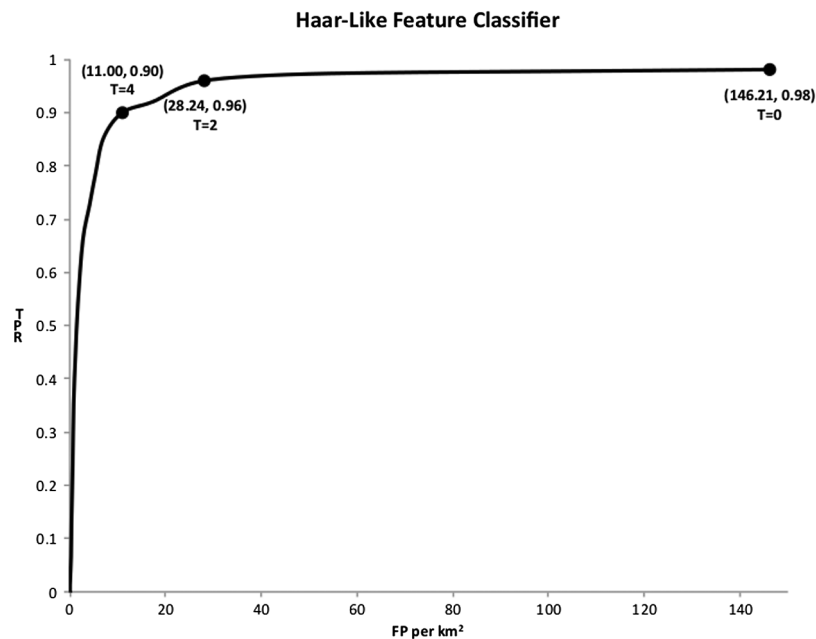


Fig. 5. Vertical axis is true positive rate (TPR) and the horizontal axis is false positives per square kilometer (FP per km²). Shows the receiver operating characteristic (ROC) curves of the Haar-like feature classifier. The three threshold points of interest are highlighted on the curve.

a way that maximizes the difference between the two conditions of positive and negative [17], [21], [22].

Choosing an optimal target prevalence is an important parameter to ensure subjects maintain focus during RSVP. When the prevalence is too high or too low, the level of engagement drops, thus increasing the chances of missing a target detection. In the work of Gerson *et al.* [15] the target prevalence was 2% for five hertz presentation speed, however experiments during development of the C3Vision system showed that a 4% prevalence of targets improves the time needed to calibrate the EEG interest classifier as well as its performance. The experiments also showed that the targets do not need to be evenly spaced, with subjects performing well even on back-to-back image chips. We use this 4% prevalence for calibration in this research, meaning in a block of 100 images only four of them are targets.

Once the EEG interest classifier is calibrated, the RSVP setup can be used to show new images to a subject and produce interest scores for each image presented. The computed EEG interest scores for each image, which is monotonically related to the probability of the MLO given the EEG, can be used to rank a set of presented images. These interest scores are not normalized and so the actual scores, which represent a high interest, will vary from subject to subject. If the EEG interest classifier has been calibrated well, images of interest will be ranked highly, providing a significant improvement in terms of time spent and detection performance when compared to traditional methods of image search [16], [17].

A. RSVP Setup

The RSVP setup that we utilized for this research is part of the C3Vision system. Before calibrating the EEG interest classifier, subjects are shown a video explaining the task and providing some information about the sonar system and the goals of the project. After the video we familiarize subjects with example

target and nontarget image chips, which are specifically not part of the calibration or testing data sets. We then conduct a few practice blocks to familiarize the subject with RSVP and correct any misconceptions regarding target identification. In this practice phase, no EEG data is collected. These steps are necessary because the subjects are not familiar with the task or the type of image.

Once a subject is confident they can differentiate targets and nontargets we begin the calibration phase to generate the EEG interest classifier. The calibration process consists of 25 blocks where each block has 100 total image chips containing four targets randomly mixed in with 96 nontargets to achieve the 4% target prevalence. The 100 image chips per block are randomly selected from the calibration pool of 240 image chips, which are 100 by 50 pixel images pulled from the same 975 image training set described in Section II for training the Haar-like feature classifier. There is no overlap in data between this calibration phase and the testing data presented in this research.

The image chips are presented to the subject at a speed of five frames per second, or five hertz. After each round the subject is presented with a graphical interface providing feedback as to the ranking assigned to the target image chips compared to their order of presentation. No subject scored below 70% on this calibration step before proceeding to the experiment, which means that the EEG interest classifier was able correctly classify the subjects neural interest 70% of the time. If a subject had scored below 70%, the calibration would have been repeated, though this was not necessary in the experiments presented in this paper.

During the actual experiment the image chips are again shown at five hertz. The experiments are split in to sections with five blocks of 100 image chips each. After each block, the subject has the power to control the start of the next block, allowing them to adjust for their own fatigue. When a subject

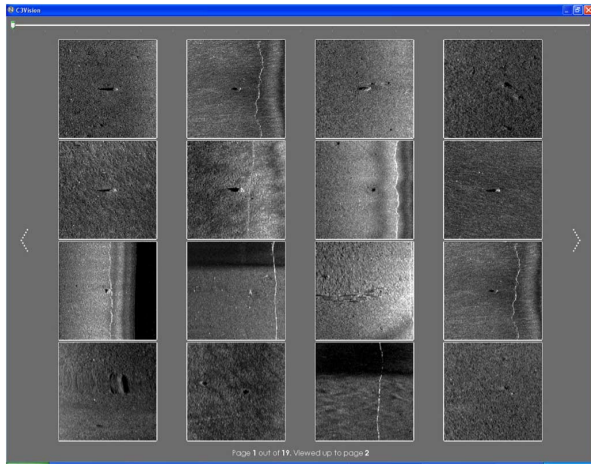


Fig. 6. Shows the interface displaying the rankings of a 500 image block. The user can move forward or backward through the rankings with the arrows or zoom to a section with the scroll bar.

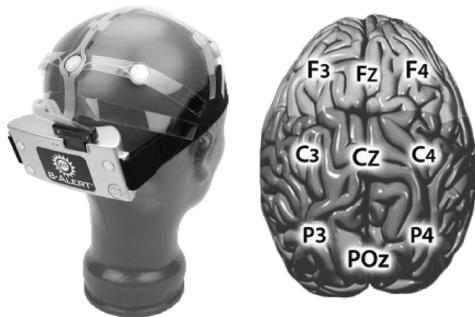


Fig. 7. Images provided by the EEG headset manufacturer, Advanced Brain Monitoring. The left image shows the B-Alert X10 wireless headset. The right image shows the layout of the sensor nodes over the brain, where the letter indicates the lobe location and the number indicates the location within a hemisphere. The lobes used by this EEG headset include the frontal (*F*), central (*C*), parietal (*P*), and occipital (*O*) where *PO* is between the parietal and the occipital. The odd numbers represent locations on the left hemisphere and even numbers represent locations on the right hemisphere.

finishes a section of 500 image chips they are all displayed in a graphical interface in order of their rankings by interest score. An example of the rankings is shown in Fig. 6. This allows the subject and the researcher to gauge the quality of the experiment as it progresses.

The EEG data collection hardware used in this experiment is the B-Alert X10 wireless headset from Advanced Brain Monitoring. This device has nine electrodes distributed over the scalp in the standardized positions of the ten-twenty electrode system [23]. Fig. 7 shows the wireless headset in the left image and the electrode layout in the right image. Each electrode placement is labeled with a letter to represent a lobe and a number to represent a location within a hemisphere. The four lobes utilized by this sensor configuration include frontal (*F*), central (*C*), parietal (*P*), and occipital (*O*), where the *PO* refers to a point between the parietal and occipital lobes. The odd numbers represent locations on the left hemisphere, while even numbers represent locations on the right hemisphere. Two reference electrodes, one behind each ear on the mastoid, are used to filter out unwanted artifacts due to muscle movement.

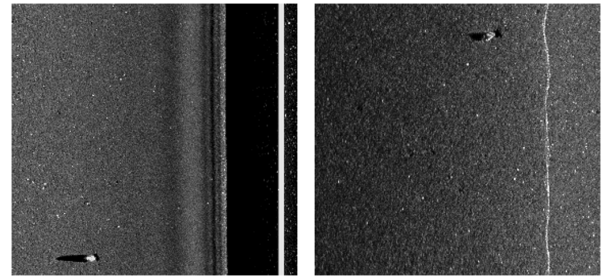


Fig. 8. Examples of the FULL data set image chips, which are approximately one-sixteenth of the original sonar image with 32 pixels of horizontal overlap and 20 pixels of vertical overlap. Both images show a mine of a different type in a very different location.

All RSVP experiments were conducted in the same office in the Department of Computer Science Building on the University of California San Diego (UCSD) campus. The subjects each sat in the same chair situated such that their heads would be approximately two feet from the computer screen when sitting up against the edge of the desk.

There were a total of 19 subjects who volunteered for the RSVP experiments from the UCSD and SSC-PAC communities. All participants had normal or corrected-to-normal vision and no history of neurological problems. There was no compensation for volunteering to be part of this experiment. Informed consent was obtained from all participants in accordance with the guidelines and approval of the UCSD Institutional Review Board.

B. Results

As a baseline for the RSVP experiments we create a preprocessed data set of image chips from the 450 sonar images in our data set. The preprocessing is very simple, splitting the sonar image in to sixteen 280 by 265 pixel regions with a 32 pixel horizontal overlap and a 20 pixel vertical overlap between regions. The result is 7200 image chips to fully represent the 450 image data set.

The RSVP experiment uses subjects with limited attention spans, and therefore 7,200 image chips is a large set of images for viewing. For this experiment, we only process a 245 subset of the images, or 3920 image chips. Because of the overlap in windows during the preprocessing, there are 168 positive image chips in this RSVP data subset, referred to in the paper as *FULL* because it fully includes the images without any true preprocessing. Fig. 8 shows two example windows from the *FULL* data set, both including a target MLO in very different locations.

The interest scores assigned to each image for a given subject are thresholded to determine a positive and negative label, and the ROC curves are created by varying the threshold over a range. Since the subjects have unique interest score domains, the range is based on the interest score data for the given subject. This same technique is used to create all the curves for RSVP results in the remainder of the paper.

Fig. 9 shows the ROC curves for the six subjects processing this *FULL* data set for this RSVP experiment. The vertical axis shows the true positive rate (TPR), which is the number of positive image chips found out of the 168 in the data set. The horizontal axis shows the number of false positives per square kilo-

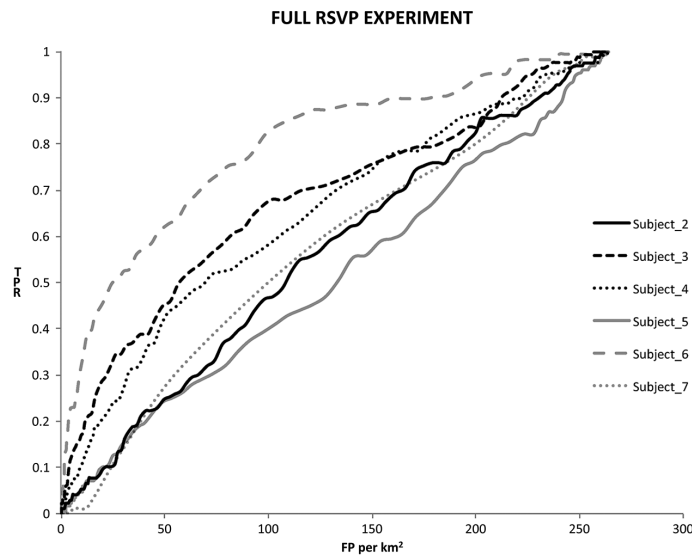


Fig. 9. Vertical axis is true positive rate (TPR) out of the 168 image chips and the horizontal axis is false positives per square kilometer (FP per km²). Shows the receiver operating characteristic (ROC) curves of the subjects processing the FULL data set.

meter (FP per km²). This is the number of image chips on average that were classified as positive but were actually negative within a square kilometer region.

The first take-away from this figure is that, overall, the six subjects perform poorly on the FULL data set. This can be attributed to the size of the image chips compared to the average size of the MLOs, which requires the subject to search the window in the short fifth of a second available. Similarly, the location of an MLO could be anywhere in the image chip because of the simple preprocessing. The overlapping setup of the FULL data set is likely to create an image chip including only a partial view of the MLO.

Another aspect of this experiment to note is the range of performance among subjects. Five of the subjects perform similarly with *Subject 6* performing substantially better. This shows that the capability of a subject to identify MLOs quickly and work well with the EEG setup is important in the results.

Overall this section shows that the RSVP experiment with the FULL data set does not perform very well, especially compared to the Haar-like feature classifier performance discussed in Subsection III-C. This leads to the main contribution of this paper, which is using the Haar-like feature classifier as the preprocessing stage to prepare image chips for RSVP. We present this capability and the experimental results in the next section.

V. COMPUTER VISION WITH RAPID SERIAL VISUAL PRESENTATION

As we have seen in the previous section, the RSVP capability does not perform well on the FULL data subset and conversely the Haar-like feature classifier performs quite well on the 450 images of our complete data set. We propose a BCI that cascades the Haar-like feature classifier before RSVP, similar to the C3Vision system. This means that the Haar-like feature classifier first processes the 450 sonar images and all positive regions become image chips that are then presented to the subjects with the C3Vision RSVP setup, which is visualized in Fig. 10. As with any two stage classifier, the second stage, which is the

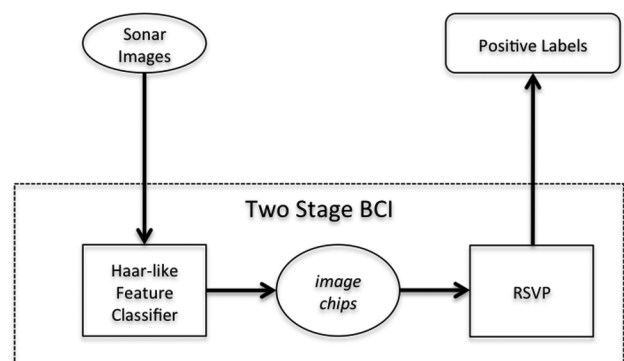


Fig. 10. Visualization of the two stage BCI, which has the Haar-like classifier as stage one and the RSVP system as stage two.

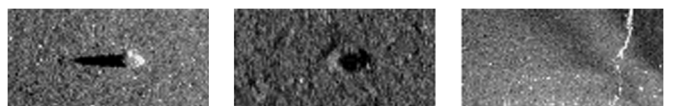


Fig. 11. Examples of the BCI image chips, including a positive, negative, and known-negative example. The known-negative is a filler used to create sparsity for the versions of the experiment that do not produce a large number of negative regions.

RSVP stage in this BCI, can only reduce the FPs and maintain the TPs at best.

The Haar-like feature classifier outputs a 79 by 29 pixel rectangle for positive regions within the processed image, which we increase to 100 by 50 pixels for the image chip. This padding around the potential positive region is meant to provide some background pixels as context with the target. Fig. 11 shows examples of positive and negative image chips with the background context padding. This figure includes a positive on the left, a negative in the middle and a known-negative on the right. In this research, the known-negatives are filler image chips to create a data set with the goal 4% prevalence of positive image chips.

TABLE I

DATA SET METRICS FOR THE THREE BCI EXPERIMENTS. THE SKIPPED COLUMNS SHOW THE TRUE POSITIVE (TP), FALSE POSITIVE (FP), AND FALSE NEGATIVE (FN) REGIONS THAT ARE SKIPPED AFTER THE FIRST STAGE HAAR-LIKE FEATURE CLASSIFIER. THE OUTPUT COLUMNS SHOW THE NUMBER OF IMAGE CHIPS OF EACH TYPE

	Skipped			Output image chips			
	TP	FP	FN	Positives	Negatives	Known-Negatives	Total
CV-0	1	64	4	146	4,238	0	4,384
CV-2	1	10	7	143	727	2,580	3,450
CV-4	1	6	16	134	281	2,985	3,400

If the padding cannot be achieved based on the location in the image then the particular region is skipped from output. Since it is not output for further processing with RSVP, it is marked as either a true positive (TP) or a false positive (FP) by comparison to ground truth. Similarly, it is possible that a MLO is missed by the Haar-like feature classifier, which is called a false negative (FN), and is thus not shown during RSVP. The total TP, FP, and FN values of the image chips not passed to RSVP system are still counted as part of the BCI classifiers overall performance.

The following subsections present three separate versions of this experiment, each using the Haar-like feature classifier with a different threshold for minimum neighbors. These correspond to the three points shown in Fig. 5, where $CV - 0$ uses a threshold of zero, $CV - 2$ uses a threshold of two, and $CV - 4$ uses a threshold of four. As the threshold increases the number of skipped TPs remains the same while the FNs increase, leading to a reduction in positive image chips produced. The number of negative image chips produced decreases greatly with the increase of the threshold. These thresholds were chosen because they cover a range of FP per km^2 while still having a TPR at or above 90%, which provides enough TPs to work with in the RSVP portion of the BCI.

Table I shows the metrics for the intermediate data sets output from each classifier version. As described, the TP, FP, and FN values that are skipped become part of the BCI's overall performance values. For CV-2 and CV-4 the number of positive and negative regions produced is so low that a large number of negative regions, which are referred to as known-negatives, are added to the RSVP data subset to keep positive image chips at approximately 4% of the total image chips. This target prevalence is essential for the performance of the RSVP system, as explained in Section IV.

A. CV-0 Experiment

The CV-0 experiment uses the Haar-like feature classifier with threshold of zero to process the 450 sonar images. This classifier version creates the largest data set for RSVP since it is the most liberal, allowing all positively labeled image chips through regardless of the classification of neighboring regions. It also has the lowest number of false negatives that are not shown during RSVP.

When processing the images, there are naturally some MLOs that are not marked as positive by the classifier. With a threshold of zero, the classifier has three such FNs that are not considered by the RSVP experiment. Also, because the image chips produced are 100 by 50 pixels to include background context, one TP and 64 FP regions are skipped and not presented during RSVP. The resulting CV-0 data set includes 4,384 image chips,

including 146 positive and 4,238 negative. All of this information is summarized in the first row of Table I.

The performance of the BCI classifier using the zero-threshold version of the Haar-like feature classifier and thirteen subjects under the RSVP process is shown in Fig. 12, with a zoomed in view near the zero-threshold point shown in Fig. 13. For each figure the vertical axis is the true positive rate (TPR), which is the number of MLO targets found by the entire BCI classifier out of the 150 targets present in the data. The horizontal axis is false positives per square kilometer (FP per km^2) based on the the known range of the images in the data set.

The full view of the curves shown in Fig. 12 shows an improvement in performance over the FULL RSVP results presented in Fig. 9. However, it is clear that the Haar-like feature classifier alone is more accurate than any subject using the CV-0 version of the BCI. Despite this less accurate performance, the zoomed in view of Fig. 13 shows some potential for the BCI classifier. Notice that many subjects are able to maintain a high TPR while reducing the FP per km^2 . Specifically, the curve created by *Subject_10* reduces the FP per km^2 by approximately 50.

B. CV-2 Experiment

The CV-2 experiment uses the Haar-like feature classifier with a threshold of two to process the 450 sonar images. It is obviously more conservative than the zero-threshold version, because two other positive images must be present for a positive label. This means, in terms of skipped images, it produces less FPs, totaling ten, and more FNs, up to seven, when compared to the CV-0 experiment in Subsection V-A. The number of TPs stays exactly the same.

There are substantial changes to the number of image chips output by the two-threshold classifier. The increase in conservativeness means that the classifier only outputs 870 total image chips, with 143 of them positive and 727 of them negative based on ground truth. Due to the nature of the RSVP experiment, the percentage of positives out of total windows is optimal around 4%, as explained in Section IV. To achieve this ratio, we add 2580 known-negatives. These are image chips that are just filler images without a target or anything similar present. An example known-negative is shown in Fig. 11 as the third image. The result of the additional images in the CV-2 data set is 3450 image chips with 143 of them positive. All of these data metrics are shown in row two of Table I.

Fig. 14 shows the ROC curves for the six BCI experiments including the two-threshold version of the Haar-like feature classifier and the RSVP process for six different subjects. There is a zoomed in view of these same curves in Fig. 15, with a focus

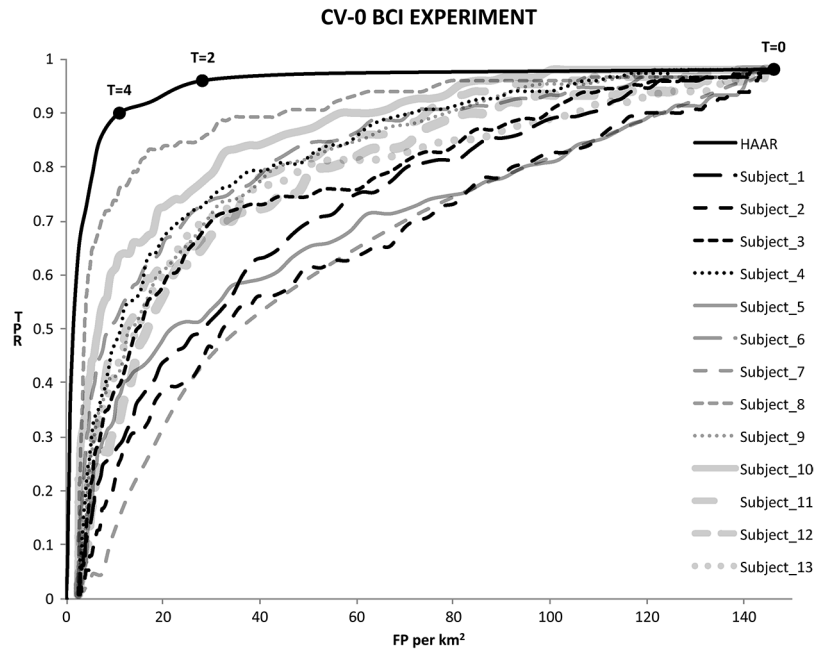


Fig. 12. Vertical axis is true positive rate (TPR) and the horizontal axis is false positives per square kilometer (FP per km²). Shows the receiver operating characteristic (ROC) curves of the BCI experiments with the Haar-like feature classifier at the zero-threshold and various subjects using the RSVP setup. The Haar-like feature classifier ROC curve, *HAAR*, is also included for comparison. The three threshold points of interest are highlighted on the curve.

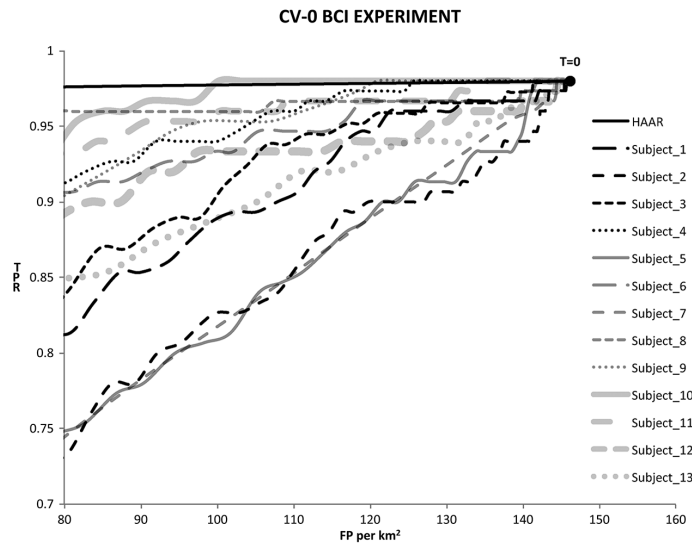


Fig. 13. Vertical axis is true positive rate (TPR) and the horizontal axis is false positives per square kilometer (FP per km²). This is zoomed in to a smaller range to show the results near the labeled threshold point of zero. Shows the receiver operating characteristic (ROC) curves of the BCI experiments with the Haar-like feature classifier at zero-threshold and various subjects using the RSVP setup. The Haar-like feature classifier ROC curve, *HAAR*, is also included for comparison.

around the $T = 2$ threshold point of the Haar-like feature classifier. For each figure the vertical axis is the true positive rate (TPR), which is the number of MLO targets found by the entire BCI classifier out of the 150 targets present in the data. The horizontal axis is false positives per square kilometer (FP per km²) based on the range of the images in the data set.

The regular view of the curves in Fig. 14 shows a BCI capability that more closely rivals the performance of the Haar-like feature classifier alone, with the BCI results for *Subject_9* outperforming the HAAR classifier. Looking at the zoomed in view of Fig. 15 we see that a couple other BCI classifiers are able to outperform the HAAR classifier at some FP per km² values.

This shows some capability improvement and great potential with higher performing subjects or with more conservative first stage classifiers.

C. CV-4 Experiment

Similar to the previous two subsections, this subsection presents the results for the BCI experiments with the first step being the Haar-like feature classifier using a threshold of four. This is the most conservative version of the classifier that we use for this research. The conservativeness results in the lowest number of skipped FPs at six and the highest number of skipped FNs, totaling sixteen. The sixteen skipped FNs means that the

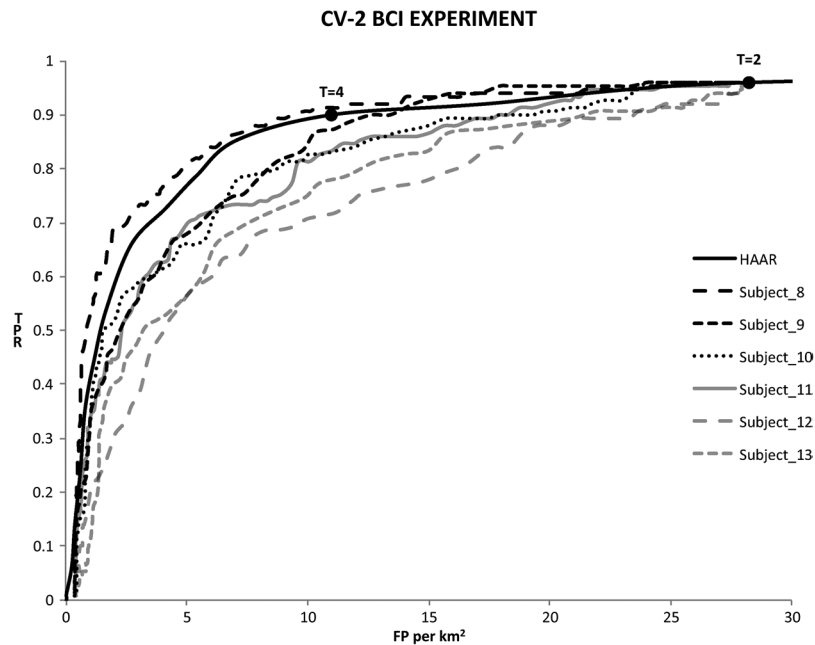


Fig. 14. Vertical axis is true positive rate (TPR) and the horizontal axis is false positives per square kilometer (FP per km^2). Shows the receiver operating characteristic (ROC) curves of the BCI experiments with the Haar-like feature classifier at the two-threshold and various subjects using the RSVP setup. The Haar-like feature classifier ROC curve, HAAR, is also included for comparison. Two of the threshold points of interest are highlighted on the curve.

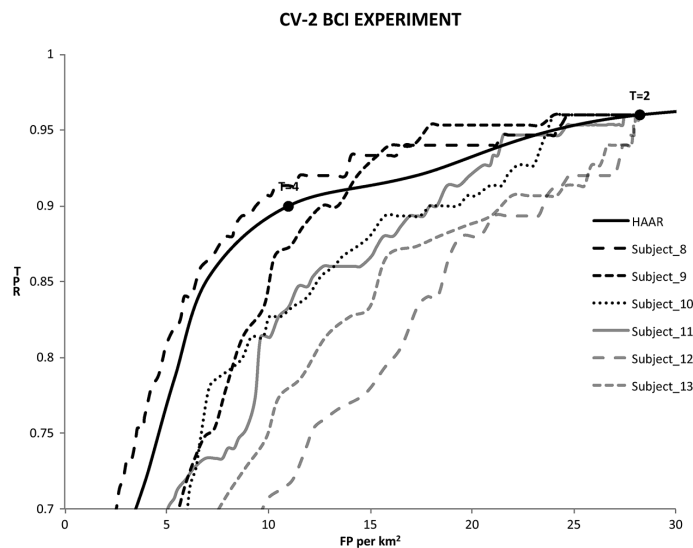


Fig. 15. Vertical axis is true positive rate (TPR) and the horizontal axis is false positives per square kilometer (FP per km^2). This is zoomed in to a smaller range to show the results near the labeled threshold point of two. Shows the receiver operating characteristic (ROC) curves of the BCI experiments with the Haar-like feature classifier at threshold two and various subjects using the RSVP setup. The Haar-like feature classifier ROC curve, HAAR, is also included for comparison.

best TPR that the BCI classifier can achieve with the RSVP step is approximately 90%.

The Haar-like feature classifier at the threshold of four produces only 415 image chips, with 134 positive and 281 negative. To achieve the 4% target prevalence, we add 2985 known-negatives to the CV-4 data set, reaching 3400 total image chips. These are the same known-negatives as discussed in Section V-B and an example is shown in Fig. 11 as the third image. All of these data metrics about the CV-4 BCI experiment and data set are shown in row three of Table I.

The five BCI experiments with the Haar-like feature classifier at a threshold of four cascaded into RSVP with five different subjects is presented in Fig. 16 The zoomed in version of the

figure, shown in Fig. 17, focuses on the threshold four point on the HAAR classifier curve. For both of these figures, the vertical axis is the true positive rate (TPR), which is again the number of MLO targets found by the entire BCI classifier out of the 150 possible targets. The horizontal axis is false positives per square kilometer (FP per km^2), which is calculated based on the known range in the sonar images.

The standard view of the curves in Fig. 16 shows a BCI capability that averages around the same performance as the Haar-like feature classifier alone over all the BCI results. The BCI results for *Subject_16* and *Subject_18* substantially outperform the HAAR classifier. Looking at the zoomed in view of Fig. 17 we see that a couple other BCI classifiers are able to

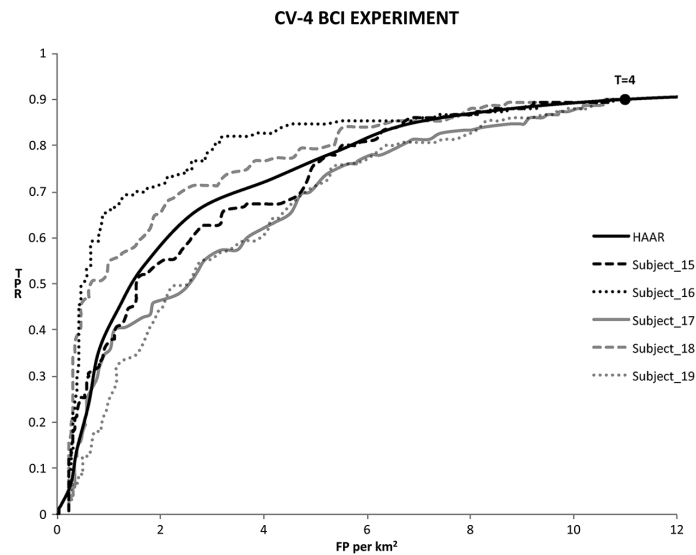


Fig. 16. Vertical axis is true positive rate (TPR) and the horizontal axis is false positives per square kilometer (FP per km^2). Shows the receiver operating characteristic (ROC) curves of the BCI experiments with the Haar-like feature classifier at threshold four and various subjects using the RSVP setup. The Haar-like feature classifier ROC curve, HAAR, is also included for comparison. The one threshold point of interest is highlighted on the curve.

marginally outperform the HAAR classifier for a range of FP per km^2 values.

This scenario with Haar-like feature classifier at threshold four provides the best performance of the three, showing that certain subjects can be used to create a BCI classifier with substantial improvements over the computer vision technique alone, but that the system is very dependent on subject capabilities at recognizing MLOs and interacting with the RSVP system.

D. Discussion

We see in this section that the two stage BCI is able to improve drastically on the FULL experiment and some subjects are able to improve over the Haar-like feature classifier. In this subsection we briefly discuss some potential reasons for these improvements.

First the improvement over the FULL experiment is relatively easy to identify. The FULL experiment divides the full images into sixteen image chips, which are large compared to the size of the targets. This makes it very difficult for the subjects to search the image and identify targets with confidence in the short time frame. All of the experiments in this section use the Haar-like feature classifier as a first stage and this produces much smaller image chips for use in the RSVP stage. These smaller image chips are much easier for the subjects to search in RSVP. It is also important to note that the experiments in this stage are much faster than the FULL experiment because of the data reduction provided by the computer vision classifier in stage one.

The increases in the threshold choice for the Haar-like feature classifier slightly reduces the TPs passed through and greatly reduces the FPs passed through to the RSVP stage. These image chips that make it through with a higher threshold are more difficult for the Haar-like feature classifier but not necessarily difficult for the human subjects. Additionally, the reduced total image chips presented to the RSVP stage causes the need for known negatives to keep a 4% target prevalence.

These known negatives are average difficulty compared to the more challenging negative image chips in the real data. This allows the subjects to better separate the positives from the negatives.

In practice, when processing data from a real mission there will be no guarantee of the 4% target prevalence used in this research. To achieve close to this goal, we would need to start with an estimate based on the history of target prevalence in real missions and the history of correctness by the Haar-like feature classifier. These combined would provide a reasonable starting point for known negative injection to achieve an estimated 4% target prevalence. When an operator processes the data, if there is an above threshold number of very strong responses, then there could be a feedback loop to rerun with a larger injection of known negative image chips. There is clearly opportunity to research this application in real scenarios in addition to the development of the algorithms themselves.

VI. BRAIN-COMPUTER INTERFACE WITH SUPPORT VECTOR MACHINE CLASSIFIER

The previous section presents a BCI that uses the Haar-like feature classifier as a first stage, cascading the output in to the RSVP process as image chips and outputting a final label for locations in the full sonar image. This has mixed success with some subjects improving the capability and some subjects performing worse than the Haar-like feature classifier alone. This section proposes training a classifier that uses the Haar-like feature from the computer vision domain with the EEG interest score feature from the human vision domain. These two feature types are combined in a feature vector to train a support vector machine (SVM) classifier, which becomes a third stage in the BCI pipeline.

Fig. 18 shows a diagram visualizing the difference between the system chain for the two stage BCI introduced in the previous section and the three stage BCI introduced in this section. The two stage BCI system uses the EEG scores to choose labels

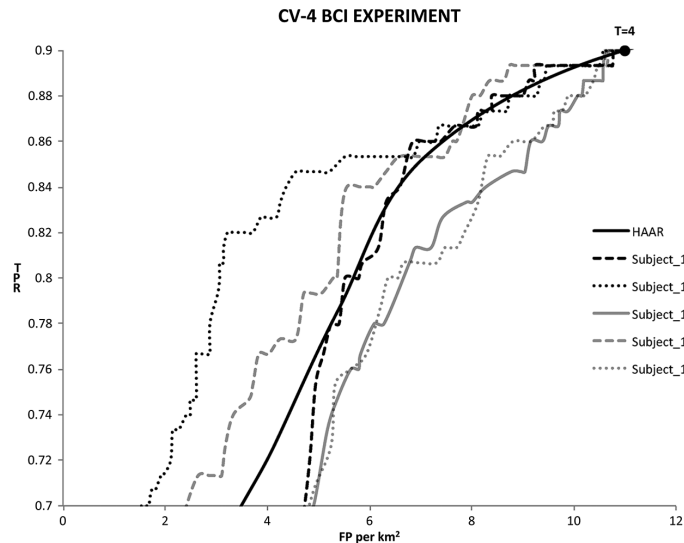


Fig. 17. Vertical axis is true positive rate (TPR) and the horizontal axis is false positives per square kilometer (FP per km²). This is zoomed in to a smaller range to show the results near the labeled threshold point of four. Shows the receiver operating characteristic (ROC) curves of the BCI experiments with the Haar-like feature classifier at threshold four and various subjects using the RSVP setup. The Haar-like feature classifier ROC curve, HAAR, is also included for comparison.

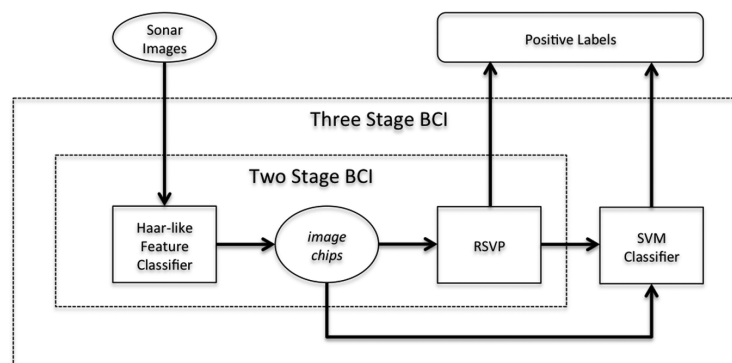


Fig. 18. Visualization of the different BCI chains introduced in this paper. The two stage BCI, introduced in Section V, has the Haar-like classifier as stage one and the RSVP system as stage three. The three stage BCI, introduced in this section, adds the third stage of an SVM classifier.

for the image chips, while the three stage BCI system passes the EEG interest scores and the image chips to the SVM classifier, which then produces the label.

To run this experiment, we divide our 450 image set in to training and testing sets of 225 images each. This way we can use the training set to create our SVM and use the testing set to compare this BCI classifier including the SVM to the Haar-like feature classifier presented in Section III and the best two stage BCI classifier presented in Section V.

We use an SVM for this part of the experiment because of its ability to handle small amounts of data for training compared to the more data hungry boosting methods. There are many training techniques that could be used for the additional classifier in the third stage of this experiment, but we leave those for future work.

An SVM is a supervised learning algorithm that, given labeled examples, outputs an optimal hyperplane to divide the examples in to positive and negative [24]. The optimal hyperplane is the one that best splits the training examples with the largest distance from the nearest example point.

Fig. 19 shows two similar visualizations of example data in multidimensional space and the optimal hyperplane selected by the SVM training. The points are shown in two dimensional space for ease of explanation. The left image shows a scenario where the data is fully separable with a hyperplane. The optimal hyperplane shown is the one that creates the largest margin, which is two times the distance from the hyperplane to the nearest examples. The right image has two additional points that create a scenario where the data cannot be fully separated. In this case the algorithm must still try to maximize the margin while at the same time minimizing the total error. The errors for the misclassified data points are labeled by ε_1 for the dark point and ε_2 for the light point.

The SVM training algorithm we utilize automatically chooses the best parameters using cross-validation, where the training data is split in to ten subsets with one for training and the remaining for testing. The training repeats for each combination of one subset for training and nine for testing to select the best parameters while finding the optimal hyperplane. We use a radial basis function (RBF) as the kernel for the SVM training

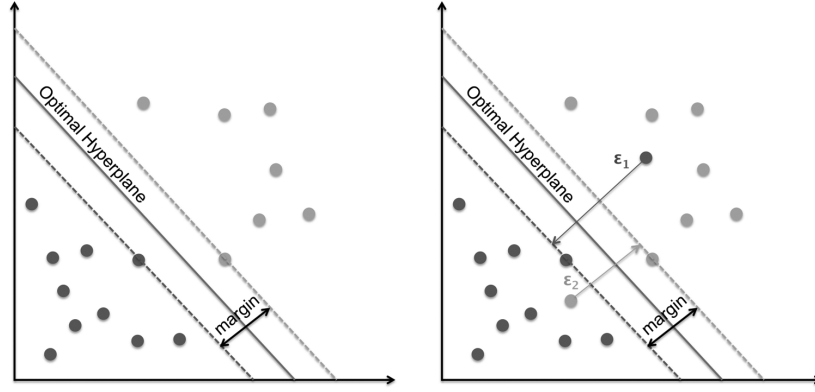


Fig. 19. Example graphs visualize the result of the SVM training in the form of an optimal hyperplane. The left image shows a fully separable data set, where the hyperplane function only maximizes the margin. The right image shows a data set that cannot be separated, where the hyperplane function maximizes the margin and minimizes the total error.

TABLE III

TESTING DATA SET METRICS FOR THE TWO BCI WITH SVM EXPERIMENTS. THE SKIPPED COLUMNS SHOW THE TRUE POSITIVE (TP), FALSE POSITIVE (FP), AND FALSE NEGATIVE (FN) IMAGES THAT ARE SKIPPED AFTER THE FIRST STAGE HAAR-LIKE FEATURE CLASSIFIER. THE OUTPUT COLUMNS SHOW THE NUMBER OF IMAGE CHIPS OF EACH TYPE THAT CASCADE TO RSVP TO PRODUCE INTEREST SCORES FOR FINAL CLASSIFICATION BY THE SVM CLASSIFIER

	Skipped			Output image chips		
	TP	FP	FN	Positives	Negatives	Total
SVM-TEST-2	0	6	4	70	375	445
SVM-TEST-4	0	4	7	68	141	209

TABLE II

TRAINING DATA SET METRICS FOR THE TWO SVM CLASSIFIERS. SHOWS THE NUMBER OF POSITIVE AND NEGATIVE IMAGE CHIPS, AS WELL AS THE TOTAL

	Positives	Negatives	Total
SVM-TRAIN-2	72	352	422
SVM-TRAIN-4	66	141	207

algorithm, which is a function that only depends on the distance from a single point such as the origin.

For this research we train SVMs on two training data sets, one composed of image chips created by the Haar-like feature classifier threshold set to two, called *SVM-TRAIN-2*. The other training set is composed of image chips created by the classifier with the threshold set to four, called *SVM-TRAIN-4*. The number of training image chips for the SVMs are shown in Table II.

When training the SVMs on the SVM-TRAIN-2 data we use a feature vector including both computer vision and human vision features. The computer vision features are the 34 Haar-like features from the Haar-like feature classifier presented in Section III and the human vision features are the six EEG interest score features corresponding to six subjects as presented in Section V. Similarly when training the SVMs on the SVM-TRAIN-4 data, the feature vector contains the same 34 Haar-like features plus the five EEG interest score features corresponding to the five subjects shown these image chips.

Once the SVMs are trained on each data set, we can test the BCI experiment with an SVM classifier as the third stage. Table III shows the testing data for the two versions with Haar-like feature classifier thresholds of two and four, called *SVM-TEST-2* and *SVM-TEST-4* respectively. In the first section we show the skipped targets that will be included in our totals after

the final classification. We see that no TPs were skipped, but a few FPs and FNs are skipped for each. The image chips that are cascaded in to the RSVP setup have similar distribution to the training data. The RSVP stage calculates interest scores, which are then used as part of the SVM classifier input to give the final label.

Fig. 20 shows the results of the BCI with SVM experiment on the SVM-TEST-2 data set created when the Haar-like feature classifier has a threshold of two. The vertical axis is the true positive rate (TPR), which is the number of MLO targets found by the entire BCI classifier out of the 74 possible targets. The horizontal axis is false positives per square kilometer (FP per km²), which is calculated based on the known range in the sonar images.

The curves in this figure are all created by testing classifiers on the SVM-TEST-2 data set. The HAAR curve uses the Haar-like feature classifier from Section III and the *Subject_8* curve uses the best BCI classifier corresponding to CV-2 from Section V. Notice that the *SVM-2 BCI* curve created by the three stage BCI outperforms both previous classifiers.

Fig. 21 shows similar results for the three stage BCI, but with Haar-like feature classifier at a threshold of four. The ROC curves are created by testing on the SVM-TEST-4 test set. Again the vertical axis is the true positive rate (TPR), which is the number of MLO targets found by they entire BCI classifier out of the possible 75 that exist. The horizontal axis is false positives per square kilometer (FP per km²), as in previous figures.

Notice the same general results where the three stage BCI using the SVM classifier outperforms the other classifiers. The HAAR curve uses the Haar-like feature classifier presented in Section III with a threshold of four and the *Subject_16* curve

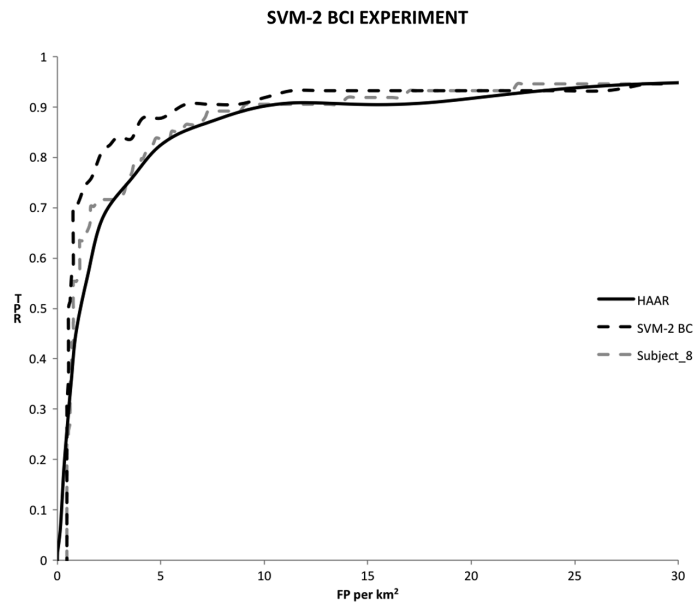


Fig. 20. Vertical axis is true positive rate (TPR) and the horizontal axis is false positives per square kilometer (FP per km²). This compares the receiver operating characteristic (ROC) curves of the three stage BCI experiment concluding with the SVM classifier, the Haar-like feature classifier, HAAR, and the two stage BCI using Subject 8, *Subject_8*. For all curves, the Haar-like feature classifier portion uses a threshold of two.

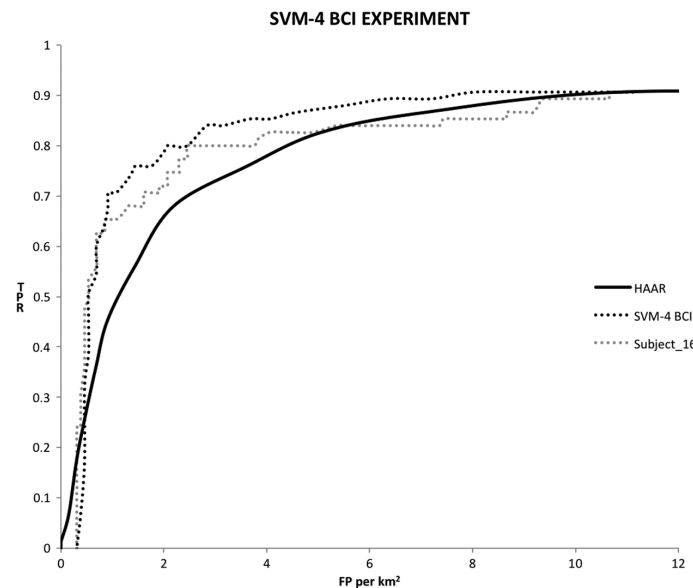


Fig. 21. Vertical axis is true positive rate (TPR) and the horizontal axis is false positives per square kilometer (FP per km²). This compares the receiver operating characteristic (ROC) curves of the three stage BCI experiment concluding with the SVM classifier, the Haar-like feature classifier, HAAR, and the two stage BCI using Subject 16, *Subject_16*. For all curves, the Haar-like feature classifier portion uses a threshold of four.

uses corresponding two stage BCI with the best performing subject. Another important element to take notice of is that the amount of data used to train these SVM classifiers is about a quarter of the data used to train the HAAR classifier, as described in Section III.

This experiment is an initial attempt at a novel concept of training a classifier using the computer vision feature and the human vision feature in the same feature vector. It shows great promise, but with limited data and breadth of experimentation, it allows space for further investigation.

VII. CONCLUSION

This paper introduces a BCI approach to the detection of mine-like objects (MLOs) in sidescan sonar imagery. The BCI system combines the complementary benefits of computer vision and human vision. We explain in depth the Haar-like feature classifier, which represents the computer vision component, and present its performance receiver operating characteristic (ROC) curve. We then provide detailed background on the rapid serial visual presentation (RSVP) process, which uses electroencephalography (EEG) based interest scores to classify images, and we present its performance ROC curve for six subjects.

The first BCI concept that we introduce uses the Haar-like feature classifier cascaded in to the RSVP process. We run experiments on this BCI system with three variations of the Haar-like feature classifier and multiple subjects per experiment. The results show that the subject processing the images and the conservativeness of the Haar-like feature classifier greatly affect the performance. In the end, we see improvement over the Haar-like feature classifier alone for some subjects and consistent improvement over RSVP classification without any preprocessing.

The second BCI concept is set up the same as the first, with an additional stage that further combines the computer vision and human vision capabilities. This third stage is a support vector machine (SVM) classifier trained on the Haar-like features and EEG interest score features. We show that this BCI system is able to provide performance improvements over the Haar-like feature classifier alone and the best two stage BCI subject performance.

This is the first use of BCI systems using EEG interest classifiers and RSVP on the problem of mine-like object detection in sidescan sonar. The combination of computer vision and human vision is a logical collaboration and we show that there is great potential for this approach to improve performance for this task.

ACKNOWLEDGMENT

The authors would like to thank P. Sajda of Columbia University and D. Rosenthal of Neuromatters, LLC for their support in the use of the C3Vision system for this research.

REFERENCES

- [1] M. Mignotte, C. Collet, P. Perez, and P. Boutheimy, "Sonar image segmentation using an unsupervised hierarchical mrf model," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1216–1231, Jul. 1998.
- [2] S. Reed, Y. Petillot, and J. Bell, "An automatic approach to the detection and extraction of mine features in sidescan sonar," *IEEE J. Ocean. Eng.*, vol. 28, no. 1, pp. 90–105, Jan. 2003.
- [3] F. Langner, C. Knauer, W. Jans, and A. Ebert, "sidescan sonar image resolution and automatic object detection, classification and identification," presented at the Proceedings. OCEANS, May 2009.
- [4] G. J. Dobeck and J. C. Hyland *et al.*, "Automated detection and classification of sea mines in sonar imagery," in *Proc. AeroSense'97. Int. Soc. Opt. Photon.*, 1997, pp. 90–110.
- [5] V. Myers and D. Williams, "Adaptive multiview target classification in synthetic aperture sonar images using a partially observable markov decision process," *IEEE J. Ocean. Eng.*, vol. 37, no. 1, pp. 45–55, Jan. 2012.
- [6] D. Williams, V. Myers, and M. Silvius, "Mine classification with imbalanced data," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 3, pp. 528–532, Jul. 2009.
- [7] J. Isaacs and J. Tucker, "Diffusion features for target specific recognition with synthetic aperture sonar raw signals and acoustic color," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2011, pp. 27–32.
- [8] P. Viola and M. Jones, "Robust real-time object detection," *Int. J. Comput. Vis.*, 2001.
- [9] S. Munder and D. M. Gavrilu, "An experimental study on pedestrian classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1863–1868, Nov. 2006.
- [10] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Proc. Int. Conf. Image Process.*, 2002, vol. 1, p. 1-900, IEEE.

- [11] M. J. Jones and D. Snow, "Pedestrian detection using boosted features over many frames," in *Proc. 19th Int. Conf. IEEE Pattern Recognit. (ICPR 2008)*, 2008, pp. 1–4.
- [12] J. Sawas and Y. Petillot, "Cascade of boosted classifiers for automatic target recognition in synthetic aperture sonar imagery," in *Proc. Meetings Acoust.*, 2013, vol. 17, p. 070074.
- [13] C. Barngrover, R. Kastner, and S. Belongie, "Semisynthetic versus real-world sonar training data for the classification of mine-like objects," *IEEE J. Ocean. Eng.*, vol. 40, no. 1, pp. 48–56, Jan. 2014.
- [14] A. Oliva, "Gist of the scene," *Neurobiol. Attention*, vol. 696, p. 64, 2005.
- [15] A. D. Gerson, L. C. Parra, and P. Sajda, "Cortically coupled computer vision for rapid image search," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 14, no. 2, pp. 174–179, Jun. 2006.
- [16] P. Sajda, E. Pohlmeier, J. Wang, B. Hanna, L. C. Parra, and S.-F. Chang, "Cortically-coupled computer vision," *Brain-Comput. Interfaces*, pp. 133–148, 2010.
- [17] P. Sajda, E. Pohlmeier, J. Wang, L. C. Parra, C. Christoforou, J. Dmochowski, B. Hanna, C. Bahlmann, M. K. Singh, and S.-F. Chang, "In a blink of an eye and a switch of a transistor: Cortically coupled computer vision," *Proc. IEEE*, vol. 98, no. 3, pp. 462–478, Mar. 2010.
- [18] N. Bigdely-Shamlo, A. Vankov, R. R. Ramirez, and S. Makeig, "Brain activity-based image classification from rapid serial visual presentation," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 16, no. 5, pp. 432–441, Oct. 2008.
- [19] A. Kapoor, P. Shenoy, and D. Tan, "Combining brain computer interfaces with vision for object categorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR 2008)*, 2008, pp. 1–8.
- [20] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *Annal. Statist.*, vol. 28, no. 2, pp. 337–407, 2000.
- [21] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, "Recipes for the linear analysis of eeg," *Neuroimage*, vol. 28, no. 2, pp. 326–341, 2005.
- [22] L. C. Parra, C. Christoforou, A. D. Gerson, M. Dyrholm, A. Luo, M. Wagner, M. G. Philiastides, and P. Sajda, "Spatiotemporal linear decoding of brain state," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 107–115, Feb. 2008.
- [23] H. H. Jasper, "The ten twenty electrode system of the international federation," *Electroencephalography Clinical Neurophysiol.*, vol. 10, pp. 371–375, 1958.
- [24] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.



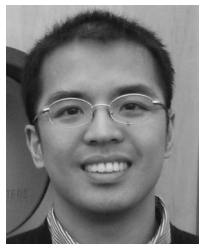
Christopher Barngrover received the B.S. degree in computer science and mathematics from Purdue University, West Lafayette, IN, USA, in 2005. He received the M.S. degree in computer science and the Ph.D. degree in computer science from the University of California San Diego, La Jolla, CA, USA, in 2010 and 2014, respectively.

He is currently a Computer Science researcher at the Space and Naval Warfare Systems Center Pacific (SSC PAC) located in Point Loma, CA, USA, where he works in the Unmanned Systems Group focusing on computer vision related aspects of robotics. He is also a Lecturer in the Department of Computer Science and Engineering at the University of California San Diego, where he teaches courses related to the software development of robotics.



Alric Althoff received the B.S. degree in cognitive science and mathematics—computer science from the University of California, San Diego (UCSD), La Jolla, CA, USA, in 2013. He is currently working toward the Ph.D. degree in computer science at UCSD.

His current research interests include hardware design for statistical signal processing, data modeling, compressed sensing, and optimization.



Paul DeGuzman received the B.S. degree in biomedical engineering from Boston University, Boston, MA, USA, in 2006, and the M.S. degree in biomedical engineering from The City College of New York, New York, NY, USA, in 2012.

He is currently a Program Manager at Neuromatters, LLC. Prior to joining Neuromatters, he was a Research Associate for the Neural Engineering Laboratory at the City College of New York, where he researched neural correlates of phantom auditory percepts. He specializes in neural signal processing, brain–computer interfaces, and psychoacoustics. He also has previous experience in medical image standardization, processing, and analysis for clinical trials.



Ryan Kastner received the B.S. degree in electrical engineering and computer engineering, and the M.S. degree in engineering from Northwestern University, Evanston, IL, USA. He received the Ph.D. degree in computer science from UCLA, Los Angeles, CA, USA.

He is currently a Professor in the Department of Computer Science and Engineering at the University of California, San Diego. He is the Codirector of the Wireless Embedded Systems Master of Advanced Studies Program. He also codirects the Engineers for Exploration Program. His current research interests reside in three areas: hardware acceleration, hardware security, and remote sensing.