Optimal Set of EEG Electrodes for Rapid Serial Visual Presentation

Kenneth E. Hild II, Santosh Mathan, Catherine Huang, Deniz Erdogmus, Misha Pavel

***Abstract*— In our application, the goal is to search through a large image to find all instances of a pre-specified, high-valued target. One approach taken to increase the throughput of this image search task is to: chop the large image into numerous small images, display them to a user at high rates one-at-a-time, and then search the simultaneously-recorded EEG data for neural activity that signifies that the user detected an instance of the target. The temporal efficiency of this EEG-based system is reduced by the overhead, which increases as the number of electrodes increases. Hence, we wish to find a minimal set of electrodes that ideally maintains the detection performance. In order to inform the design of future EEG-based image search systems, in this paper we find the 12 out of 32/64 most important electrodes for detection using 5 different feature selection methods. They optimal set includes all 5 occipital and the 2 most frontal electrodes.**

I. INTRODUCTION

Imagine that you wish to find, as quickly as possible, all instances of a high-valued target in a large image. The first approach is to load the image onto a screen and then have a human pan around and zoom in and out of the image. A second approach is to use a computer algorithm to automatically search for targets in the image. The former is problematic because it is relatively slow. The latter is problematic because the detection performance of automatic search algorithms (with the constraint that there are either no or few false positives) is oftentimes relatively low. Fortunately, a third approach has recently become a reality [1].

The third approach for searching images is an EEG-based system that incorporates a human and a computer algorithm. In this approach, the human detects targets and the computer algorithm detects precisely *when* the human detects targets. As it turns out, humans are very good at detecting the presence of a target within an image even if they view the image for as little as 50 or 100 ms. Based on this fact, the images in an EEG-based system are chopped into smaller images (so the need for zooming is removed), which are then rapidly displayed one-at-a-time (so the need for panning is removed) on a screen directly in front of the (human) user. This image presentation paradigm is known as rapid serial visual presentation (RSVP). If we stopped here, the EEG-based system would encounter a problem. Even though humans can detect targets shown for a fraction of a second, we are not very good at signifying precisely when we detect targets, which becomes a problem as we increase the rate at which images are shown. More specifically, the standard deviation of the time between image onset and the behavioral response that signifies detection (e.g., a button press) can be larger than the image duration. This is where the computer algorithm comes in. Even though computer algorithms are not very good at detecting arbitrarily-shaped targets (e.g., in images), they do a relatively good job at detecting approximately fixed-shaped neural signatures, such as the one that occurs in EEG data when the user detects a target. Moreover, since this particular neural signature (known as the P300 [2]) is approximately time-locked to the image onset, the standard deviation of the time between image onset and the neural signature is relatively small.

The reason one might elect to use an EEG-based system is because it is more temporally efficient than manual (human) search, as has been shown by several different research groups [3],[4]. The efficiency calculation oftentimes, however, does not include the overhead associated with setup or tear down. The setup overhead includes the time required to prepare the scalp, place the electrodes, and check the impedances of the electrodes. The tear down overhead includes the time required to remove and clean the electrodes and to clean the user's scalp. In our experience, the total overhead associated with a 32-electrode EEG system is on the order of 30 minutes. Let's assume that the overhead for a 32-electrode system is 30 minutes and that 20 minutes of this total scales proportionally to the number of electrodes. Based on this information, the overhead for a 12 and a 128-electrode system is $10 + 20 * (12/32) \approx 18$ and 90 minutes, respectively. Let us further assume that the overhead can be ignored if it constitutes less than or equal to 10% of the time spent viewing images in a single session. In this case, the overhead is negligible only if the user views images for a minimum of 3, 5, and 15 hours per session when using a 12, 32, and 128-electrode system, respectively.

Increasing the number of electrodes is detrimental to the temporal efficiency of an EEG-based system, but it can allow an increase in the detection performance. In our experience, as the number of electrodes is increased the performance increases at an exponentially-decreasing rate. More specifically, we expect a relatively large improvement if we change the number of electrodes from 1 to 3 and

This work was supported by Defense Advanced Research Projects Agency contract #NBCHC080030. The views, opinions, and/or findings contained in this presentation are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency (DARPA) or the Department of Defense (DOD).

K.E. Hild, C. Huang, M. Pavel are with the Dept. of Biomedical Engineering, Oregon Health & Science University, Portland, OR 97239, USA k.hild@ieee.org, huang@csee.ogi.edu, pavel@bme.ogi.edu

D. Erdogmus is with the Dept. of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA derdogmus@ieee.org

S. Mathan is with Human Centered Systems at Honeywell Laboratories, Redmond, WA 98052, USA santosh.mathan@honeywell.com

TABLE I
LIST OF METHODS

#	Description	Abbreviation
1.	Ratio of target to distractor energy	Erg
2.	Mutual information between data and class label	MI
3.	Principle feature analysis	PFA
4.	Forward selection	FS
5.	Sequential forward floating selection	SFFS

we expect a relatively small improvement if we change the number of electrodes from 64 to 128. The final number of electrodes used in an EEG-based system must represent a balance between the opposing forces of temporal efficiency and detection performance.

In this paper, we assume that the optimal number of electrodes is 12 and our goal is to find the optimal location for these 12 electrodes. In an attempt to find the 12 best locations, we apply 5 different electrode selection methods to 5 different datasets, each of which was collected using either 32 or 64 electrodes.

II. ELECTRODE SELECTION METHODS

We can consider the temporal signal produced at a given electrode (in response to a single image) as a single feature. Hence, we can use any feature selection algorithm to select electrodes. The selection methods we use, which are listed in Table I, are meant to be a representative set of the vast array of methods that have been described in the literature. The 5 methods include 3 that are generative and 2 that are discriminative, which means that a classifier is not required for the optimization of the former and is required for the latter. The Erg, MI, and PFA methods are generative and the FS and SFFS methods are discriminative.

Feature selection is a specific type of feature reduction. The most common feature reduction methods include principle components analysis (PCA) and linear discriminant analysis (LDA). We do not include either of these 2 methods because they, as do many other feature reduction methods, learn a linear transformation from the original set of features to the set of lower-dimensional output features. In other words, they require access to the full set of original features, whereas our goal is to reduce the number of features (electrodes) that must be collected.

In the paragraphs that follow, we describe the 5 methods for selecting electrodes. In this discussion, we use $x_{k,m,n}$ to denote the EEG data collected at electrode k, image m, and time n, where each image has an associated temporal window (e.g., the temporal window might extend from 100 ms to 500 ms after the onset of its associated image). In addition, we use $x^T_{k,m,n}$ and $x^D_{k,m,n}$ to denote the data that correspond to a target image and a distractor (non-target) image, respectively, whereas the lack of a superscript indicates that either we do not know or do not care if the given image is a target or a distractor.

A. Ratio of Target to Distractor Energy

The first electrode selection method, Erg, is based on energy. More specifically, this approach selects the electrodes that have the largest values of target to distractor energy, where the energy of each is found by averaging over images and time. The Erg performance metric for the k^{th} electrode is given by,

$$J_k = \frac{\frac{1}{NM^T}\sum_{m=1}^{M^T}\sum_{n=1}^{N}(x^T_{k,m,n})^2}{\frac{1}{NM^D}\sum_{m=1}^{M^D}\sum_{n=1}^{N}(x^D_{k,m,n})^2}, \quad (1)$$

where $n = 1$ corresponds to the first sample in the temporal window, N is the total number of samples in the temporal window, and M^T, M^D are the total number of target and distractor images, respectively.

B. Mutual Information

The second electrode selection method, MI, is based on mutual information. More specifically, this approach selects the electrodes that have the maximum mutual information between the data and the class label. This approach involves 12 iterations. In each iteration, one previously-unselected electrode is added to the set of selected electrodes, $\mathcal{K}$ (which is initialized to $\emptyset$). Hence, the cardinality of $\mathcal{K}$ increases by one each iteration and once an electrode is added to $\mathcal{K}$, it cannot be removed. The mutual information between the data and the class label is given by $I(x_{k\in\mathcal{K},m,1:N}, C_m)$, where $x_{k\in\mathcal{K},m,1:N}$ denotes the $N|\mathcal{K}|$ samples of the temporal window associated with image m and the electrodes in $\mathcal{K}$, $|\mathcal{K}|$ is the cardinality of $\mathcal{K}$, C_m is the class label of image m, $C_m \in \{T, D\}$, T is the class label used for targets, and D is the class label used for distractors. We approximate the mutual information by expressing it as the Renyi marginal entropy minus the Renyi conditional entropy, as described in a previous publication [5]. Hence, in each iteration we can express the MI performance metric for set $\mathcal{K}$ as,

$$J_{\mathcal{K}} = H_2(x_{k\in\mathcal{K},m,1:N}) - H_2(x_{k\in\mathcal{K},m,1:N}|C_m) \quad (2)$$

where $H_2(x)$, which represents Renyi's quadratic entropy of x, is estimated using Parzen windows. The approach used here was obtained by taking the previously-published feature reduction method [5] and making modifications so that it can be used for feature selection [5].

C. Principle Feature Analysis

The third electrode selection method, PFA, is based on a modification of PCA published by Lu et al. [6]. More specifically, this approach forms a matrix whose columns consist of the dominant eigenvectors of the PCA solution, clusters the rows using K-means, finds for each cluster the index of the row that is closest to its corresponding mean, and returns the set of indices (one for each cluster). The set of returned indices represents the electrodes that, according to PFA, should be kept. The PFA performance metric for set $\mathcal{K}$ is given by,

$$J_{\mathcal{K}} = f_{K\text{-}means}([q_1\, q_2\, \cdots\, q_{12}]) \quad (3)$$

where $f_{K\text{–}means}(Q)$ denotes that K-means is applied to the rows of matrix Q and q_i represents the i^{th} dominant eigenvector (the eigenvector having the i^{th} largest eigenvalue).

There have been other attempts to modify PCA so that it can be used to select features. These methods commonly select features by finding the index of each dominant eigenvector that has the maximum magnitude. However, this type of approach can lead to highly-correlated output features, a problem which is avoided by PFA.

D. Forward Selection

The fourth electrode selection method, FS, is a greedy method that selects the electrodes that have the maximum classification accuracy. As with MI, this approach involves 12 iterations. In each iteration, one previously-unselected electrode is added to the set of selected electrodes, $\mathcal{K}$ (which is initialized to $\emptyset$). Unlike MI, the selection in each iteration of FS is based on the classification accuracy of a Bayes-MOG classifier using 10-fold cross validation. The classification accuracy is measured by estimating the receiver operating characteristic (ROC) curve and finding the area under this curve (AUC). In each iteration, the FS performance metric is given by,

$$J_{\mathcal{K}} = \frac{1}{10}\sum_{i=1}^{10} f_{AUC}\left(f_{Bayes\text{–}MOG}(x_{k\in\mathcal{K},m\in\mathcal{M},1:N}, C_{m\in\mathcal{M}})\right) \tag{4}$$

where 10 is the number of cross validation sets, $f_{Bayes\text{–}MOG}(x_{k\in\mathcal{K},m\in\mathcal{M},1:N}, C_{m\in\mathcal{M}_i})$ denotes that the Bayes-MOG classifier is trained on the EEG data corresponding to the images in set $\mathcal{M}_i$ using the data from electrodes in $\mathcal{K}$, $f_{Bayes\text{–}MOG}()$ returns the posterior probability that each of the images in the disjoint test set (all m s.t. $m \notin \mathcal{M}_i$) contains a target, and $f_{AUC}()$ computes the AUC.

E. Sequential Forward Floating Selection

The fifth electrode selection method, SFFS, selects the electrodes that have the maximum classification accuracy [7]. Like both MI and FS, this is an interative method. However, unlike MI and FS, at each iteration SFFS can either add or remove electrodes to the existing set, $calK$, and SFFS iterates until it reaches a local optimum. As is the case for FS, the selection in each iteration is based on the AUC of a Bayes-MOG classifier using 10-fold cross validation. In each iteration, the SFFS performance metric is given by,

$$J_{\mathcal{K}} = \frac{1}{10}\sum_{i=1}^{10} f_{AUC}\left(f_{Bayes\text{–}MOG}(x_{k\in\mathcal{K},m\in\mathcal{M},1:N}, C_{m\in\mathcal{M}})\right) \tag{5}$$

which is identical to (4) since the difference between the 2 methods is only how $\mathcal{K}$ is updated in each iteration.

III. COMPARISON

The 5 feature selection methods are compared using 5 different datasets, which are briefly described in Table II. All 5 datasets were collected using an EEG system manufactured by Biosemi (Amsterdam, Netherlands) and they were collected at a sample rate of 256 Hz using either 32 or 64 electrodes. The electrode locations used in all 5 datasets are listed in Table III (we discard the 32 extra electrodes from each of the 64-electrode datasets). The locations are given in spherical coordinates. The first column represents the degrees of inclination from Cz (where positive values correspond to the right hemisphere) and the second column represents the degrees of azimuth (from T7 for the left hemisphere and from T8 for the right hemisphere, positive values correspond to anti-clockwise rotations). The pre-processing consists of filtering the data using a sixth-order Butterworth band pass filter with cutoff frequencies of 1 and 40 Hz.

To compare the methods we do the following. First, we find, for each electrode selection method, the optimal set of 12 electrodes for each subject of each of the first 4 datasets (producing a total of $6 + 2 + 10 + 10 = 28$ results per method). Second, we combine the 28 results in order to produce a single optimal set, which we refer to as the overall-optimal set, of 12 electrodes per electrode selection method (the process of combining results across subjects/datasets is discussed below). Third, we test each solution using the fifth dataset (where we use a Bayes-MOG classifier and 10-fold cross validation).

As stated above, separate performance metric results are obtained for each subject of each dataset. For the Erg method, it is trivial to combine these results across multiple subjects/datasets since the performance metric is computed for each electrode separately. For the other 4 methods, on the other hand, the performance metric is only computed for sets of electrodes. The set of optimal electrodes is expected to differ, if only slightly, from one subject/dataset to the next and it is not feasible to compute the performance metric for all possible combinations. Hence, to combine results across subjects/datasets for MI, PFA, FS, and SFFS, we measure (for each method separately) the percentage of time each electrode is included in one of the 28 optimal sets.

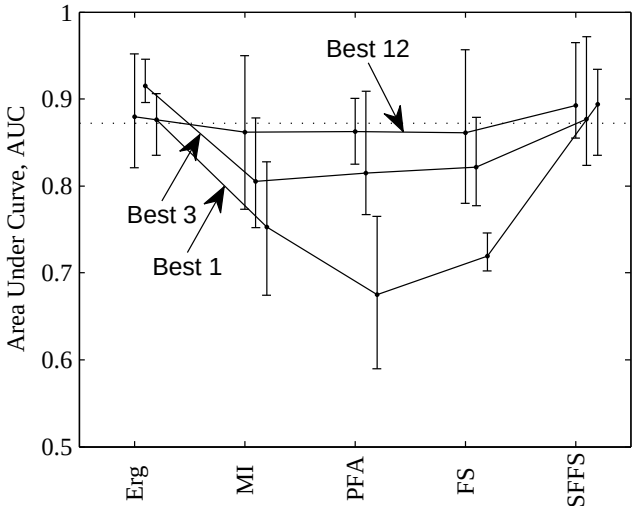

Fig. 1. AUC for all 5 electrode selection methods for 1, 3, 12, and 32 electrodes.

TABLE II

LIST OF DATASETS

Dataset	Subjects	Image Duration(s) in (ms)	Target	Masks
1.	6	300	POL	No
2.	2	60, 100	SAM sites	No
3.	10	50, 100, 150, 200	SAM sites	Yes
4.	10	50, 100, 150, 200	SAM sites	Yes
5.	3	100, 150	SAM sites	No

TABLE III

LIST OF ELECTRODES

Electrode	Name	Location		Electrode	Name	Location	
1.	Fp1	-92	-72	17.	O2	92	-72
2.	AF3	-74	-65	18.	PO4	74	-65
3.	F7	-92	-36	19.	P4	60	-51
4.	F3	-60	-51	20.	P8	92	-36
5.	FC1	-32	-45	21.	CP6	72	-21
6.	FC5	-72	-21	22.	CP2	32	-45
7.	T7	-92	0	23.	C4	46	0
8.	C3	-46	0	24.	T8	92	0
9.	CP1	-32	45	25.	FC6	72	21
10.	CP5	-72	21	26.	FC2	32	45
11.	P7	-92	36	27.	F4	60	51
12.	P3	-60	51	28.	F8	92	36
13.	Pz	46	-90	29.	AF4	74	65
14.	PO3	-74	65	30.	Fp2	92	72
15.	O1	-92	72	31.	Fz	46	90
16.	Oz	92	-90	32.	Cz	0	0

IV. RESULTS

Fig. 1 shows the AUC (averaged over 10 cross validation sets and the 3 subjects of the fifth dataset) for the overall-optimal set of 12 electrodes for each method. The error bars indicate the minimum and maximum values (over the 3 subjects). The dashed line in this figure is the mean AUC when all 32 electrodes are used. For sake of comparison, we also include results for the overall-optimal set of 1 and 3 electrodes. All 5 methods perform nearly the same when each uses its overall-optimal set of 12 electrodes. If we consider the performance of 1, 3, and 12-electrode solutions then, for our data, the SFFS and Erg methods perform noticeably better than the MI and FS methods and much better than the PFA method.

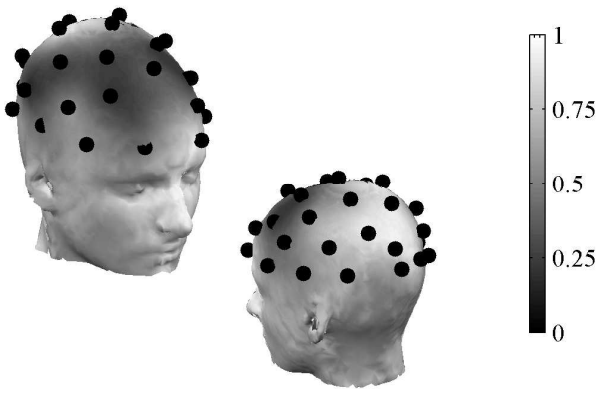

Fig. 2. Scalp plot where larger values correspond to increased importance.

TABLE IV

LIST OF 12 OPTIMAL ELECTRODES FOR SFFS

Electrode	Name	% Optimal	Electrode	Name	% Optimal
15	O1	0.56	16	Oz	0.47
11	P7	0.55	30	Fp2	0.47
12	P3	0.51	18	PO4	0.45
1	Fp1	0.50	17	O2	0.44
32	Cz	0.49	20	P8	0.44
14	PO3	0.48	3	F7	0.42

Fig. 2 shows a scalp plot corresponding to the SFFS method. Larger values (lighter shades of gray) indicate that the electrode in question is considered more important. To improve visualization, we normalized the results in this figure so that they span the range of 0 to 1. Notice that the results are spatially smooth even though we made no attempt to enforce spatial smoothness. Table IV shows the overall-optimal set of 12 electrodes for the SFFS method. Also shown in this table are the percentage of times each of the 12 appeared in the 28 optimal solutions corresponding to all combinations of subjects/datasets. For sake of comparison, the worst electrode appears in 15% of the 28 optimal solutions.

V. DISCUSSION

The Erg method is neither discriminative (whereas the FS and SFFS methods are) nor does it take into account what electrodes were selected previously (whereas all 4 other methods do, albeit in a greedy fashion). Nevertheless, it performs nearly as well as SFFS.

The results in Fig. 1 indicate that it may be possible to reduce the number of electrodes even further without a large loss in detection performance. However, keep in mind that it is not uncommon for one or more electrodes to be dropped during a recording session (an electrode is "dropped" when its impedance becomes unacceptably high, which renders the associated signal unusable). Using a larger number of electrodes provides robustness against dropped electrodes.

REFERENCES

[1] L. Parra, C. Christoforou, A. Gerson, M. Dyrholm, A. Luo, M. Wagner, M. Philiastides, P. Sajda, "Spatiotemporal linear decoding of brain state," *IEEE Signal Proc. Magazine*, Vol. 25, pp. 107-115, Jan. 2008.

[2] T.W. Picton, "The P300 wave of the human eventrelated potential," *Journal of Clinical Neurophysiology*, Vol. 9, No. 4, pp. 456-479, 1992.

[3] Y. Huang, D. Erdogmus, S. Mathan, M. Pavel, "A Fusion Approach for Image Triage using Single Trial ERP Detection," Intl. EMBS Conference on Neural Engineering (CNE '07), Kohala, Hawaii, pp. 473-476, May 2007.

[4] A.D. Gerson, L.C. Parra, P. Sajda, "Cortically coupled computer vision for rapid image search," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, Vol. 14, No. 2, pp. 174-179, 2006.

[5] K.E. Hild II, D. Erdogmus, K. Torkkola, J.C. Principe, "Feature extraction using information-theoretic learning," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 9, pp. 1385-1392, Sept. 2006.

[6] Y. Lu, I. Cohen, X.S. Zhou, Q. Tian, "Feature selection using principal feature analysis," Intl. Conf. on Multimedia, Augsburg, Germany, pp. 301-304, Sept. 2007.

[7] P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, Vol. 15, No. 11, pp. 1119-1125, Nov. 1994.