Decision-Level Fusion of EEG and Pupil Features for Single-Trial Visual Detection Analysis

Ming Qian*, Mario Aguilar, *Member, IEEE*, Karen N. Zachery, Claudio Privitera, Stanley Klein, Thom Carney, and Loren W. Nolte, *Life Senior Member, IEEE*

***Abstract*—Several recent studies have reported success in applying EEG-based signal analysis to achieve accurate single-trial classification of responses to visual target detection. Pupil responses are proposed as a complementary modality that can support improved accuracy of single-trial signal analysis. We develop a pupillary response feature-extraction and -selection procedure that helps to improve the classification performance of a system based only on EEG signal analysis. We apply a two-level linear classifier to obtain cognitive-task-related analysis of EEG and pupil responses. The classification results based on the two modalities are then fused at the decision level. Here, the goal is to support increased classification confidence through the inherent modality complementarities. The fusion results show significant improvement over classification performance based on a single modality.**

***Index Terms*—Brain–machine interface, decision fusion, EEG, image triage, machine learning, pupillary response, rapid serial visual presentation (RSVP), single-trial detection.**

I. Introduction

In VISUAL image classification applications, human visual analysts are substantially more effective than computer visual systems at parsing a scene and recognizing target objects. The main challenge is that we usually have a limited number of qualified human visual analysts and, in most applications, an enormous amount of visual images need to be classified.

As an example of the demand on resources, it was estimated that in 2004 alone, nearly 40 million mammograms were taken in the USA [4]. The number of available expert radiologists falls well short of such overwhelming demands. To alleviate these problems, effective image classification techniques are needed to rapidly screen the high volume of imagery and sort out a much smaller subset of images that merits further detailed analysis.

Manuscript received April 27, 2008; revised December 4, 2008. First published March 27, 2009; current version published June 12, 2009. This works was supported by the Defense Advanced Research Projects Agency (DARPA) under Contract HM1582-05-C-0041 and by the Neurotechnology for Intelligence Analysts. *Asterisk indicates corresponding author*.

*M. Qian is with Teledyne Scientific and Imaging LLC, Research Triangle Laboratory, Durham, NC 27713 USA (e-mail: mqian@teledyne.com).

M. Aguilar and K. N. Zachery are with Teledyne Scientific and Imaging LLC, Research Triangle Laboratory, Durham, NC 27713 USA (e-mail: mdaguilar@teledyne.com; kzachery@teledyne.com).

C. Privetera, S. Klein, and T. Carney are with the Vision Processing Laboratory, University of California, Berkeley, CA 94720 USA (e-mail: claudiop@berkeley.edu; sklein@berkeley.edu; neurometrics@speakeasy.net).

L. W. Nolte is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708 USA (e-mail: lwn@ee.duke.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2009.2016670

EEG is a psychophysiological modality we can use as a basis for single-trial image classification tasks. In previous studies [7] and [15], it was found that a steorotypical spatiotemporal EEG response would be triggered when a target image was shown among a sequence of nontarget/distractor images. Single-trial analysis using simple classifiers can be applied to detect the spatiotemporal EEG signature. Prototype EEG-based systems [8], [13], [18] have shown significant improvements over a traditional manual search approach in terms of processing throughput, speed, and accuracy in high-volume visual classification tasks.

The pupil dilation reflex is mediated by inhibition of the parasympathetic Edinger–Westphal oculomotor complex and sympathetic activity. It has long been documented that emotional and sensory events elicit a pupillary reflex dilation [10], [11], [20]. Task-evoked pupil dilation was found to be a function of the cognitive workload/attention required to perform the task [2], [3]. Experiments by Privitera *et al.* [16] found a significant pupil response for visual target detection events.

In this paper, we develop a pupillary response feature-construction and -selection procedure to extract the useful features that perform best under a linear discriminant analysis (LDA) classifier. The LDA classifier decodes these features to achieve analysis of target detection event. To leverage EEG signals, we apply a two-level LDA classification scheme [18] on cognitive task-related EEG responses. While the EEG-based method captures the evolving patterns of brain dynamics, the pupillary response provides us a direct indicator of emotional arousal or alertness triggered by the visual detection sensory stimulus.

While the pupil dilation reflex is controlled by the peripheral nervous system, the brain activity measured by the EEG signals reflects the central nervous system. Therefore, the simultaneous use of these modalities offers potential complementarities that can serve to improve classification performance. Our experimental results demonstrate that the decision-level fused classification achieves significant improvement over the classification performance based on a single modality (EEG or pupillary response).

II. Methods

A. Participants

To compare the increased accuracy and efficiency of the proposed method, five participants (one female and four males) participated and received no payment. All had normal or corrected to normal vision and reported no history of neurological

0018-9294/$25.00 © 2009 IEEE

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on December 3, 2009 at 19:30 from IEEE Xplore. Restrictions apply.

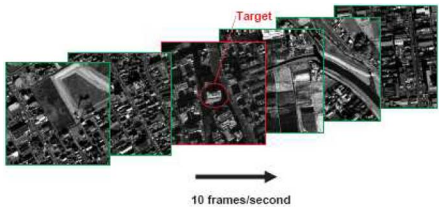


Fig. 1. RSVP sequence (Imagery Credit DigitalGlobe).

problems. All participants were image analysts with more than five years of experience.

B. Behavioral Paradigm

During the visual target detection task, participants were presented with a sequence of urban scene images. Each experimental run consisted of the presentation of thousands of images utilizing the rapid serial visual presentation (RSVP) paradigm. The sequence consisted of a few targets interspersed among nontarget or distractor images. For every experimental session, each subject went through several runs of 60 s each for a total time of 5–6 min on average. Subjects were given a 1-min break after each session to relax their eyes. Images were presented at 10 Hz (ten images per second) where a continuous sequence of images is presented in the center of the screen. To help the subjects engage the task consistently, they were instructed to push a button as soon as they detected a target image.

C. Stimuli

Image processing software was used to segment wide-area urban scene images into 500 × 500 pixel image chips. Target images were selected that contained targets centered on the images (Fig. 1). The target-to-distractor ratios ranged from 1% to 5%. These ratios reflect the natural target densities of urban imagery and the selected target detection tasks.

D. EEG Signal Acquisition

EEG signal was acquired in a sound isolating chamber using a 64-channel Biosemi EEG system (http://www.biosemi.com) while subjects performed the visual detection tasks. EEG was recorded continuously with a sampling rate of 512 Hz. Electrodes were placed according to the international 10–20 EEG placement standard.

E. Pupil Signal Acquisition

Pupil data were acquired using an EyeLink1000 video-based eye tracking device (http://www.sr-research.com) that offers a 1000 Hz sampling rate using an infrared camera. The right eye was used for each subject. Pupil data recording was synchronized with EEG signal acquisition and resampled to 512 Hz.

F. EEG Signal Preprocessing

Signal preprocessing procedures were applied to the recorded EEG signal (see Figs. 2 and 3). DC baseline drifts and abrupt voltage jumps caused by muscle movements were removed by a second-order Butterworth bandpass filter.

At the beginning of the experiment, participants were asked to blink their eyes multiple times. The recorded EEG signal during this eye blink period can be analyzed to determine the linear component associated with eye blinks [14]. The eye blink artifact can be filtered out of the recorded EEG during the RSVP task using the derived eye blink component.

After the eye blink artifact removal, the average voltage amplitude of all 64 signal channels was used as the single reference signal and this reference signal was subtracted from the signal channels in order to remove the unnecessary noise in the "reference free" data that were acquired by active electrodes [6].

The preprocessed EEG signal was segmented into 600-ms epochs time-locked on visual stimulus onset events for each target/nontarget/distractor trial. The segmented epochs were used as inputs for single-trial classification. To reduce the classification computational load, the data in each epoch were down-sampled from 512 to 120 Hz.

G. Pupil Signal Preprocessing

The continuous pupil size data can be corrupted by eye blinks or head movements. The EyeLink1000 monitoring system has a signal loss detection mechanism implemented to identify all corrupted signal regions. Since most blinks lasted very short period of time (average blink time is around 90 ms) compared to the valid pupil response time period (average pupil response lasts more than 1 s) toward a visual event, we interpolated the continuous data to fill in the missing data segments. A moving average filter was used to smooth the pupil area data to improve the SNR.

The preprocessed pupil data were segmented into 1550-ms epochs time-locked on visual stimulus onset events for each trial. A single-trial epoch starts at 300 ms before the stimulus onset and ends at 1250 ms after the stimulus onset (Fig. 5). The segmented epochs were used for feature extraction and single-trial classification. Since pupil dilation is a relatively slow response, we can reduce the computation time for feature extraction and classification by further down-sampling the pupil data from 512 to 20 Hz. Consequently, each single-trial pupil data epoch had 32 pupil diameter sample points.

H. Pupil Dilation Feature Construction

To characterize the pupillary response more thoroughly and improve the SNR, we identified a set of features related to the pupillary response and derived them from the original pupil diameter data (e.g., the average velocity and acceleration of the pupil size change illustrated in Fig. 4).

To extract pupil dilation features, the pupil diameter, pupil diameter rate of change, acceleration, and pupil area are derived for all time samples inside the 1550-ms pupil epoch window. The epoch can be further divided into prestimulus, after-stimulus, and "critical" periods (Fig. 5).

We identified a total of 22 features that could be derived from raw pupil dilation data (Table I). Some of them are general

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on December 3, 2009 at 19:30 from IEEE Xplore. Restrictions apply.

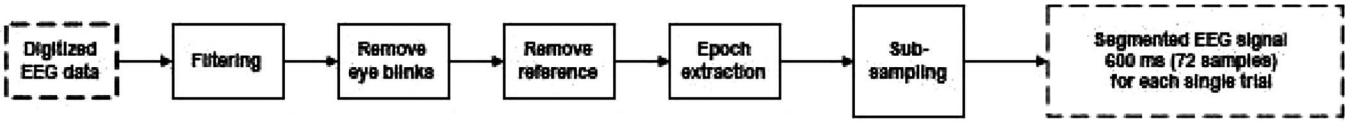


Fig. 2. EEG preprocessing procedure.

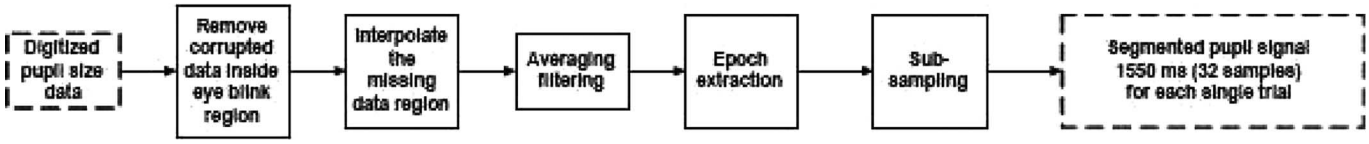


Fig. 3. Pupil signal preprocessing procedure.

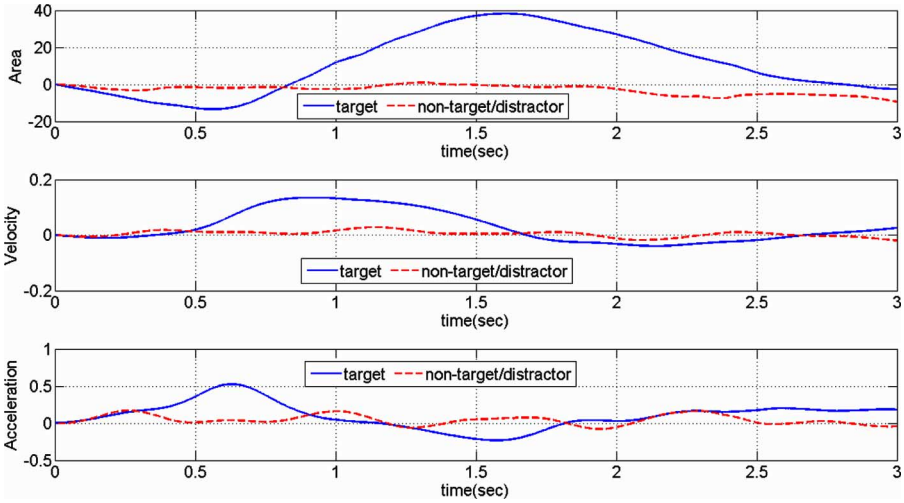


Fig. 4. Comparison of the average pupil area/velocity/acceleration change time course (target versus nontarget/distractor) for one subject. The velocity/acceleration features show earlier differentiations compared to the area feature. We achieve more thorough characterization by deriving velocity/acceleration features.

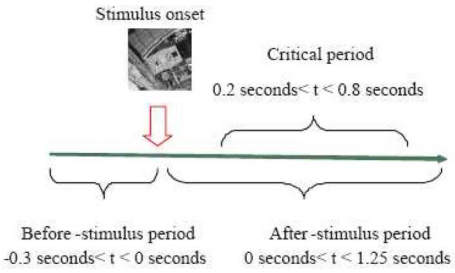


Fig. 5. Time windows defined for pupil dilation feature extraction: before-stimulus time period (-0.3 s $< t < 0$ s); after-stimulus time period (0 s $< t <$ 1.25 s); "critical" time period (0.2 s $< t <$ 0.8 s) is regarded as the most likely time period during which the pupillary response is most salient (Imagery Credit DigitalGlobe).

features such as maximum dilation velocity while the others were studied in [1] and [16].

I. Pupil Feature Selection

The benefits of feature selection are twofold: it reduces the computational cost of classification by reducing the number of features that need to be calculated and improves classification accuracy by fitting a simpler model based on finite number of training samples. Given the 22 candidate features described in Table I, a subset of d features, $d < 22$, is selected that performs the best under the specific classifier we choose to use (e.g., a linear discriminant classifier).

A large number of algorithms have been proposed for feature subset selection [9]. The optimal methods (e.g., exhaustive search method) are usually too time-consuming even for a moderate number of features. Many suboptimal methods have been developed to strike a balance of computational efficiency and optimality.

Sequential forward selection (SFS) method starts from an empty feature subset and sequentially selects the next most significant feature and adds it to the feature subset to maximize the cost function J until a predefined feature number is obtained [21]. Sequential backward selection (SBS) starts from a subset with all d features and iteratively deletes the least significant feature until a predefined feature number is obtained [12]. Both SFS and SBS methods have the so-called nesting problem: once a feature is added/deleted, it cannot be deleted/added anymore. Sequential float feature selection (SFFS) method avoids nesting problem by correcting earlier "mistakes" by backtracking: first enlarges the feature subset by adding l most significant features using SFS, then deletes r least significant features using

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on December 3, 2009 at 19:30 from IEEE Xplore. Restrictions apply.

TABLE I
PUPILLARY FEATURES CONSTRUCTED AS CANDIDATES FOR FEATURE SELECTION

1	maximum accumulated dilation velocity measured in the T_{after}
2	maximum accumulated dilation velocity difference in the T_{before} and the T_{after}
3	maximum dilation velocity in the T_{after}
4	maximum dilation acceleration in the T_{after}
5	maximum velocity in the T_{after} - mean velocity in the T_{before}
6	maximum velocity difference in the T_{after} and in the T_{before}
7	maximum area in the $T_{critical}$ - mean area before $T_{critical}$
8	maximum pupil velocity in $T_{critical}$
9	product of feature 7 and 8
10	maximum diameter in the T_{after}
11	maximum diameter in the T_{after} - mean diameter in the T_{before}
12	absolute difference between maximum normalized (4-sample) moving average diameter in the T_{after} and T_{before}
13	difference between maximum normalized (5-sample) diameter deviations from moving average in the T_{after} and T_{before}
14	difference between maximum normalized (5-sample) diameter deviations from moving average in the T_{after} and T_{before}
15	difference between maximum normalized (5-sample) diameter gradient from moving average in the T_{after} and T_{before}
16	difference between maximum normalized absolute value slopes (using a 4-sample interval) in the *Tafter* and *Tbefore*
17	difference between maximum normalized absolute value slopes (using a 4-sample interval) in the T_{after} and T_{before}
18	difference between maximum normalized diameter gradient moving sums in the T_{after} and T_{before}
19	percentage difference of samples in the T_{after} and T_{before} that have diameters larger than the diameter at the stimulus onset
20	average energy in the T_{after} and T_{before}
21	power ratio of the peak samples in a 16-sample window centered around the stimulus onset versus all the samples in the window
22	difference between contour energy in the T_{after} and T_{before}

SBS. And the l and r are determined dynamically ("floating") so that the optimal solution can be approximated [17].

To assure the robustness of our feature-selection algorithm, we randomly permuted the training samples after calculating their 22 candidate features; then we used average tenfold cross-validation classification performance as our feature-selection cost function. In our experiments, for each subject, we selected the best features based on the training data and then these selected features are used for testing.

J. EEG-Based Classification

By linearly combining EEG channels, an aggregate representation of the data can be obtained [15]. Let $\mathbf{x}(t)$ be the observed vector of multidimensional EEG amplitude reading (after signal preprocessing) at time t, an optimal projection weighting vector $\mathbf{w}_{\mathrm{amp}}$ can be derived based on a training set and so that a 1-D projection $y_{\mathrm{amp}}(t)$ can be generated:

$$y_{\mathrm{amp}}(t) = \mathbf{w}_{\mathrm{amp}}^T \mathbf{x}(t) = \sum_{i=1}^{N} w_{\mathrm{amp}_i} x_i(t) \tag{1}$$

where N is the number of EEG electrodes (channels). For the visual detection task, we can find an optimal projection weight matrix $\mathbf{w}_{\mathrm{amp}}$, which can generate a $y_{\mathrm{amp}}(t)$ that maximally discriminates at time t corresponding to target and nontarget/distractor trial condition.

A more robust output can be generated by averaging over the T samples in the short temporal window $\overline{y}_{\mathrm{amp}} = T^{-1} \sum_{t=1}^{T} y_{\mathrm{amp}}(t)$. The hypotheses are $\mathbf{H_0}$ (nontarget/distractor) and $\mathbf{H_1}$ (target). The posterior likelihood that a trial belongs to the target class is assumed to follow a logistic function distribution $p(\mathbf{H}_1 | y_{\mathrm{amp}}(t)) = 1/(1 + e^{-\overline{y}_{\mathrm{amp}}})$ [14].

In this paper, LDA classifiers are applied on eight overlapping 50-ms temporal windows with onset time ranging from 150 to 500 ms following the stimulus onset in step of 50 ms.

While LDA classifiers can be applied over multiple short temporal windows after stimulus onset, the likelihood output values of these first-level LDA classifiers (called local classifier) can be used as inputs to a feature-level classifier (called global classifier) to capture the linear pattern across multiple local windows, as illustrated in Fig. 6 [7], [15].

By linearly combining local classifier outputs, an aggregate representation of global pattern can be obtained. Let $\mathbf{y}$ be the observed vector of local classifier linear projection outputs, a weighting vector $\mathbf{w}_{\mathrm{window}}$ can be derived based on training data to generate a 1-D projection z

$$z = \mathbf{w}_{\mathrm{window}}^T \mathbf{y} = \sum_{i=1}^{M} w_{\mathrm{window}_i} y_i \tag{2}$$

where M is the number of local classifiers. The projection z serves as an estimate of global pattern. The likelihood of a trial belonging to the target class is assumed to follow a logistic function distribution $p(\mathbf{H_1}|z) = 1/(1 + e^{-z})$.

Receiver operating characteristic (ROC) curves can be obtained using $p(\mathbf{H_1}|z)$ and compare it with a threshold η, which can take on values ranged in [0 1]. The decision rule is

$$\begin{cases} p(\mathbf{H_1}|z) \geq \eta, & u = 1 \\ p(\mathbf{H_1}|z) < \eta, & u = 0 \end{cases}$$

or vice versa, where $u = 1$ represents the classifier's decision to declare a target and $u = 0$ represents the classifier's decision to declare a nontarget/distractor.

K. Pupillary-Based Classification

By linearly combining multiple pupillary-based features, an aggregate representation of the features can be obtained. Let $\mathbf{d}$ be the observed vector of the selected pupillary features, an optimal projection weighting vector $\mathbf{w}_{\mathrm{pupil}}$ can be derived based on a training set and so that a 1-D projection y_{pupil} can be generated:

$$y_{\mathrm{pupil}}(t) = \mathbf{w}_{\mathrm{pupil}}^T \mathbf{d} = \sum_{i=1}^{D} w_{\mathrm{pupil}} d_i \tag{3}$$

where D is the number of pupillary features selected using the SFFS method. The projection $y_{\mathrm{pupil}}(t)$ can be assumed to follow some distributions of the exponential family and is regarded as a better estimate of neurophysiologic activity than any individual pupillary response feature.

ROC curves can be obtained using $p(\mathbf{H_1}|y_{\mathrm{pupil}})$ and comparing it with a threshold θ. θ can take on values within the range [0, 1]. The decision rule can be

$$\begin{cases} p(\mathbf{H_1}|y_{\mathrm{pupil}}) \geq \theta \\ p(\mathbf{H_1}|y_{\mathrm{pupil}}) < \theta \end{cases}$$

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on December 3, 2009 at 19:30 from IEEE Xplore. Restrictions apply.

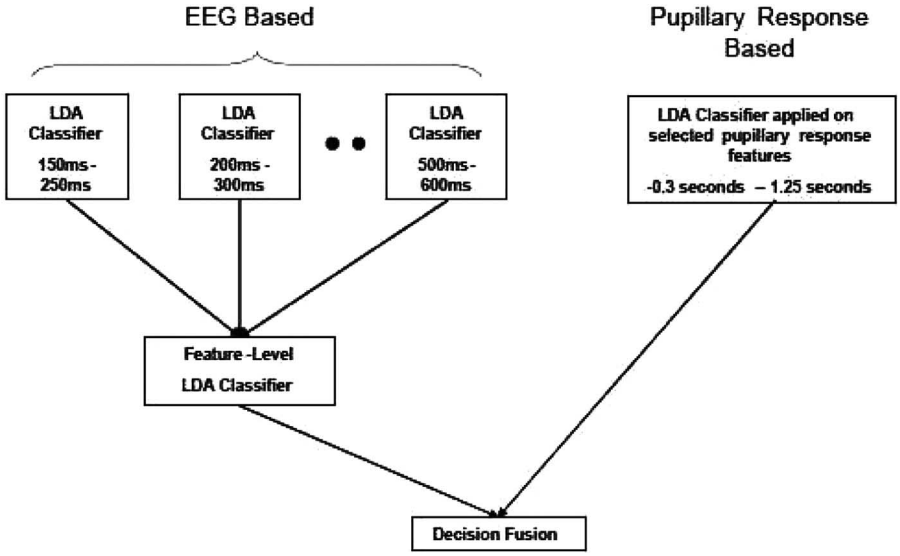


Fig. 6. Fusion scheme: classifier results are derived using EEG and pupil signals; thereafter they are fused at the decision level.

or vice versa where $u_{\text{pupil}} = 1$ represents the classifier's decision to declare a target and $u_{\text{pupil}} = 0$ represents the classifier's decision to declare a nontarget/distractor.

L. Decision Fusion

It has been demonstrated that decisions from multiple classifiers (e.g., EEG and pupillary response) can be combined to improve the overall classification performance by exploring the complementarities of source and pattern of two modalities [5]. For example, while the EEG signal can be obscured by muscle movement artifact, pupillary modality is more robust to this effect. Here, the goal is to enhance classification confidence through the fusion of the two decisions.

Decision-level fusion optimally fuses two decisions according to the operating points on their ROC curves at which each decision was made with certain probability of detection (P_d) and probability of false alarm (P_f). Individual detectors (EEG- and pupil-based) receive as inputs the N-dimensional EEG observation vector $\mathbf{x}$ and the D-dimensional pupillary feature observation vector D, and output the decisions using the LDA classifiers. The decisions drawn by individual classifiers are denoted as u_k, for $k = 1$ (EEG based) and $k = 2$ (pupillary response based), where $u_k = 0$ if kth classifier's decision is $\mathbf{H_0}$ and $u_k = 1$ if kth classifier's decision is $\mathbf{H_1}$. Individual classifier's decision u_k depends only on the observation vectors $\mathbf{x}$ and $\mathbf{p}$:

$$u_1 = \alpha(\mathbf{x}) = \begin{cases} 0, & \text{EEG - based classifier decides } \mathbf{H_0} \\ 1, & \text{EEG - based classifier decides } \mathbf{H_1} \end{cases} \tag{4}$$

$$u_2 = \beta(\mathbf{p}) = \begin{cases} 0, & \text{pupil - based classifier decides } \mathbf{H_0} \\ 1, & \text{pupil - based classifier decides } \mathbf{H_1}. \end{cases} \tag{5}$$

The performance characteristics of classifier k can be specified by $P(u_k|\mathbf{H_j})$, where $P(u_k = 1|\mathbf{H_0}) = P_{fk} =$ the probability of false alarm (false positives) and $P(u_k = 1|\mathbf{H_1}) = P_{dk} =$ the probability of detection (true positives). Using these probabilities, the likelihood ratio value of a binary decision variable has a simple form as

$$\lambda_{\text{decision}}(u_k) = \frac{P_r(u_k|\mathbf{H_1})}{P_r(u_k|\mathbf{H_0})} = \begin{cases} \dfrac{P_{dk}}{P_{fk}}, & \text{if } u_k = 1 \\ \dfrac{(1-P_{dk})}{(1-P_{fk})}, & \text{if } u_k = 0. \end{cases} \tag{6}$$

The decision at the fused level depends only on local decisions and their probability of detection P_{dk} and probability of false alarm P_{fk}. Since individual classifiers are based on different modalities, the simplest assumption is that these two decisions are statistically independent. The fusion likelihood ratio

$$\lambda_{\text{fusion}}(u_1, u_2) = \frac{P(u_1, u_2|\mathbf{H_1})}{P(u_1, u_2|\mathbf{H_0})} = \prod_{k=1}^{2} \lambda(u_i) = \prod_{k=1}^{2} \frac{P(u_k|\mathbf{H_1})}{P(u_k|\mathbf{H_0})} = \prod_{k=1}^{2} \left(\frac{P_{dk}}{P_{fk}}\right)^{u_k} \prod_{k=1}^{2} \left(\frac{1-P_{dk}}{1-P_{fk}}\right)^{1-u_k}. \tag{7}$$

The optimal decision fusion rule uses the fusion likelihood ratio as a classification decision variable and then compares it to threshold β for decision u [19]:

$$u = F(u_1, u_2) = \begin{cases} 1, & \text{if } \lambda_{\text{fusion}}(u_1, u_2) \geq \beta \\ 0, & \text{if } \lambda_{\text{fusion}}(u_1, u_2) < \beta. \end{cases} \tag{8}$$

Fig. 7 illustrates the fusion of two operating points from two separate ROC curves. By varying the threshold β value (corresponding to different decision rules), five operating points can be derived for the fused ROC curve. The probability of detection and false alarm for decision-fused classifier operating points can be calculated based on (9) and (10):

$$P_{d\,\text{fusion}}(\beta) = \sum_{\lambda_{\text{fusion}} \geq \beta} P(\lambda = \lambda_{\text{fusion}}|\mathbf{H_1}) \tag{9}$$

$$P_{f\,\text{fusion}}(\beta) = \sum_{\lambda_{\text{fusion}} \geq \beta} P(\lambda = \lambda_{\text{fusion}}|\mathbf{H_0}). \tag{10}$$

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on December 3, 2009 at 19:30 from IEEE Xplore. Restrictions apply.

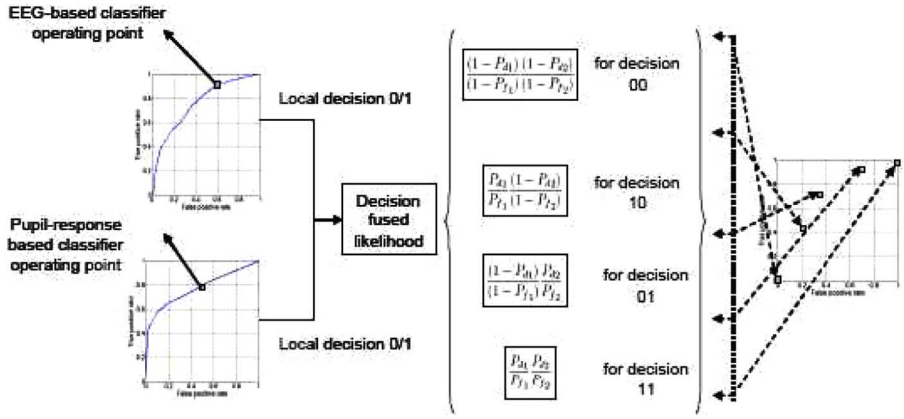


Fig. 7. Fusion of two operating points from two separate ROC curves, five fused operating points can be generated depending on where the threshold is set against the four fused likelihood ratio values.

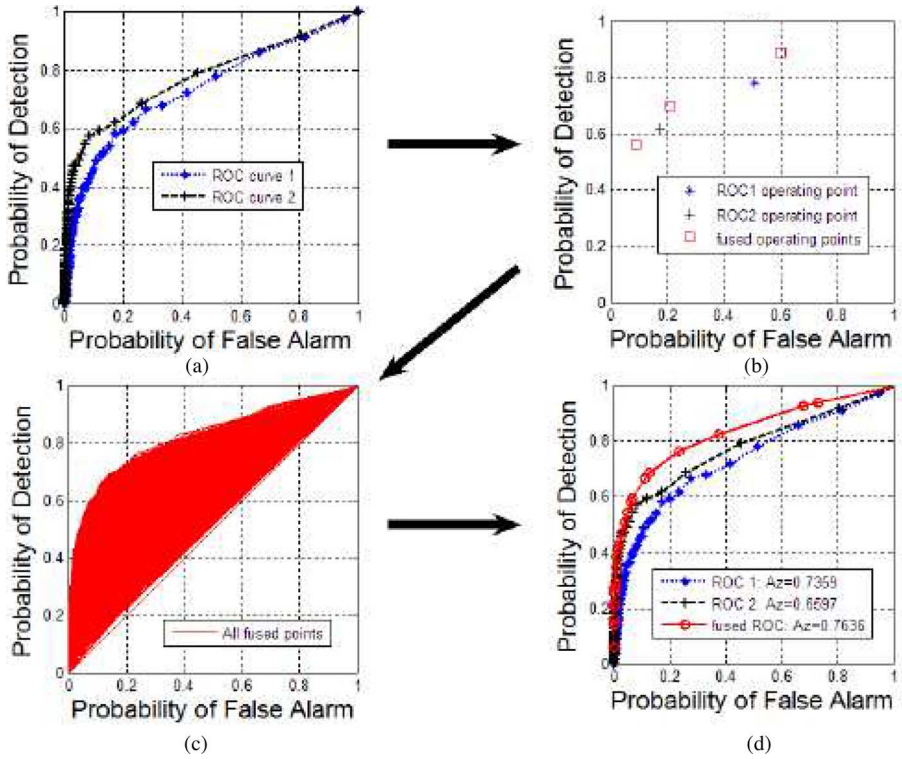


Fig. 8. Illustration of fusing two ROC curves. (a) Multiple operating points on each ROC curve. (b) Five fused operating points can be generated by fusing one operating point from one ROC curve and another operating point from the other ROC curve. (c) Large number of fused operating points can be generated by exhausting all combinations of individual ROC curves' operating points. (d) Convex hull of all the fused operating points can be used to represent the optimum improved ROC curve for decision-fused results.

The entire decision-fused ROC curve can be obtained by deriving a series of $P_{d\,\mathrm{fusion}}(\beta)$ and $P_{f\,\mathrm{fusion}}(\beta)$ using multiple combinations of operating points on EEG-based and pupillary-response-based classifier ROC curves. Fig. 8 illustrates the process of generating a fused ROC curve by fusing two ROC curves. Each of the individual ROC curve has multiple operating points with different P_d and P_f [shown in Fig. 8(a)]. Every time, one operating point from one ROC curve can be fused with one operating point from another ROC curve, as illustrated in Fig. 8(b). Since each ROC curve can have multiple (e.g., 100) operating points, a large number of fused operating points [as illustrated in Fig. 8(c)] can be generated to exhaust all operating point combinations. To evaluate the optimum improvement achievable by the decision-level fusion, the convex hull of these points [as illustrated in Fig. 8(d)) is used as the fused ROC curve to represent the best possible area under ROC curve (A_z) for the fused result.

In practice, the decisions made by the EEG-based and pupillary-response-based classifiers should be compared with the actual ground truth using a training set so that $P(u_1, u_2|\mathbf{H_0})$, $P(u_1, u_2|\mathbf{H_1})$, and $\lambda_{\mathrm{fusion}}(u_1, u_2)$ can be evaluated empirically. Using the training set, we can choose the optimal operating

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on December 3, 2009 at 19:30 from IEEE Xplore. Restrictions apply.

TABLE II
NUMBER OF TARGET AND NONTARGET/DISTRACTOR IMAGES, TARGET VERSUS NONTARGET/DISTRACTOR RATIOS IN TEST SETS

Subject ID	Test set	Target image	NonTarget image	ratio
1	1	150	3341	1:22
	2	27	3464	1:128
2	1	150	3341	1:22
	2	27	3437	1:127
3	1	150	3341	1:22
	2	27	3437	1:127
4	1	156	3277	1:21
	2	27	3464	1:128
5	1	156	3121	1:120
	2	27	3374	1;125

points associated with the EEG-based classifier, the pupillary-response-based classifier, and the decision-level classifier, respectively.

III. RESULTS

A. Training and Test Runs

In the training set used for all five subjects, we manipulated the target–distractor ratio to 10% to obtain sufficient target samples for training the classifiers. Data collected using this imagery serve to train the EEG-based classifier, the pupil-based classifier, selecting the relevant pupil features and training the decision fusion classifier.

Classification performance was evaluated based on two test sets with different target density per sequence (see Table II).

B. Feature Selection and Classification Results

Different feature subsets are selected for five subjects based on the training set. Among them, features 1, 2, 6, 7, and 11 (Table I) are selected for all five subjects. The EEG-based, the pupillary-response-based, and the decision fusion classifier performance were evaluated using Az as the metric. The results for subjects 1–5 over two test sets are shown in Fig. 9.

For the pupillary-response-based classification, the classification performance on test sets with lower target density per sequence is consistently better than the classification performance on test sets with higher target density per sequence across all five subjects. This result is consistent with the stimulus probability effect: a less frequently presented stimulus elicits a stronger pupillary response while a more frequently presented stimulus elicits a weaker pupillary response and is also reported by Privitera *et al.* [16].

The results shown in the figures are analyzed for significance using a T-test method (Tables III and IV). The inverse hyperbolic arctangent rescaled A_z was used, because improvement on the end of higher A_z value should be regarded as having higher significance. The null hypothesis (mean of single-modality performance is equal to mean of the fused performance) is tested against the alternative hypothesis (mean of single-modality performance is less than the fused performance). For test set 1, the null hypothesis was rejected at significance level of 0.05.

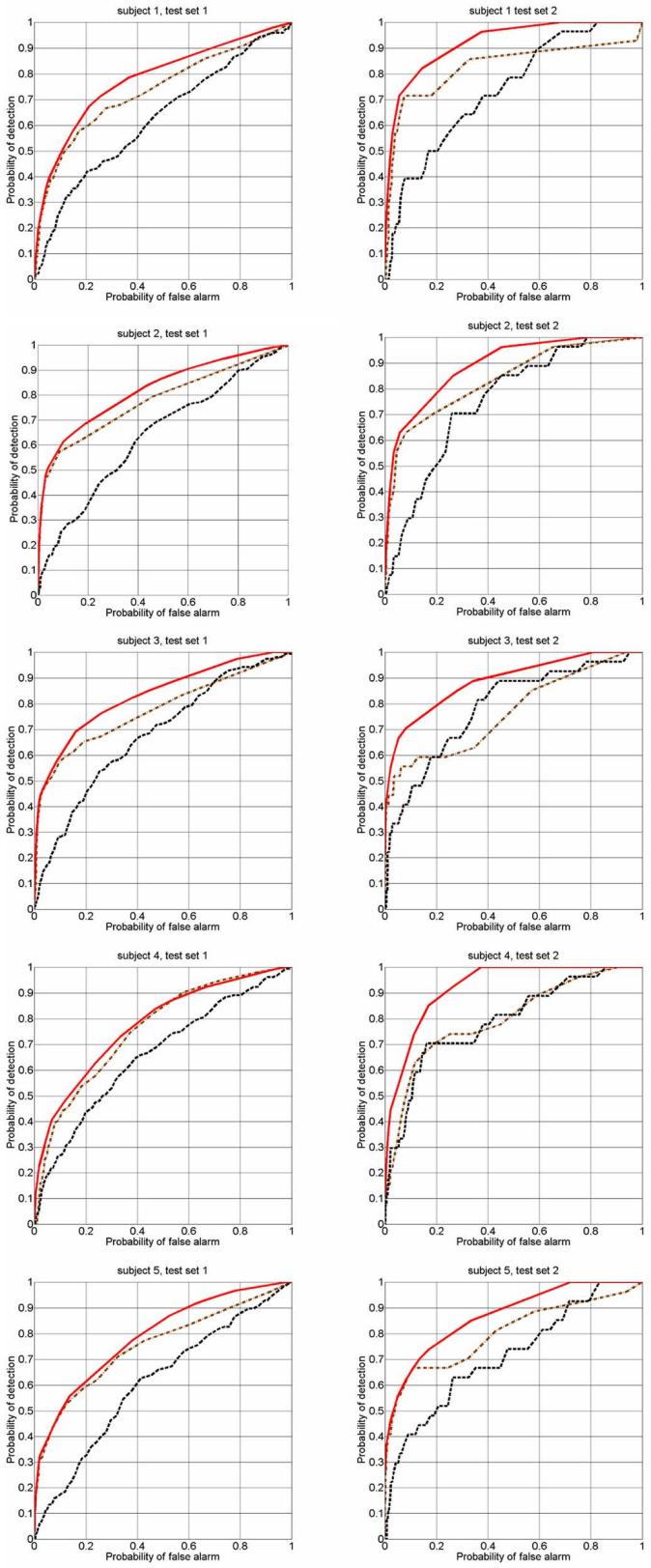


Fig. 9. Classification results for five subjects on test sets 1 and 2: dash-dotted curves are the ROC curves for the EEG-based results; dashed curves are the ROC curves for the pupillary-response-based results; solid curves are the ROC curves for the fused results.

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on December 3, 2009 at 19:30 from IEEE Xplore. Restrictions apply.

TABLE III
PERFORMANCE MEASURES USING THE INVERSE HYPERBOLIC ARCTANGENT A_z RESCALED ON TEST SET 1

Subject ID	EEG based	Pupil based	Fusion
1	0.9415	0.7319	1.0464
2	1.0428	0.756	1.1623
3	1.0342	0.8345	1.1995
4	0.9974	0.7965	1.0352
5	0.7543	0.611	1.0733

TABLE IV
PERFORMANCE MEASURES USING THE INVERSE HYPERBOLIC ARCTANGENT RESCALED A_z ON TEST SET 2

Subject ID	EEG based	Pupil based	Fusion
1	1.1837	0.9472	1.6076
2	1.2094	0.9843	1.4307
3	1.0125	1.0881	1.4214
4	1.0964	1.107	1.6049
5	1.1067	0.9245	1.336

The p-value is 0.0124. For test set 2, the null was rejected at significance level of 0.05. The p-value is 0.00026.

IV. DISCUSSION

The goal of our methodology is to increase both the throughput of imagery and overall accuracy of the assessment. Cognitive task-related pupillary response was shown to provide a complementary modality for single-trial image classification tasks. We developed a pupillary response feature-extraction and -selection procedure to extract features that perform best under a linear analysis classifier. We also applied a two-level linear analysis classification scheme on cognitive task-related EEG responses. The classification results based on both modalities are further fused at the decision level. The fusion results show significant improvement over classification performance using a single modality.

For pupil-based classification, a set of 22 pupil dilation features were evaluated in this study. Features 1, 2, 6, 7, and 11 (Table I) were selected by SFFS algorithm for all five subjects. The implication is that the combination of simple measures of velocity change, pupil area change, and diameter change, triggered by the target events, are the most effective.

We have demonstrated that the resulting image classification paradigm can be used to improve the accuracy of visual target search on high-volume imagery. As the signals our methodology exploit are not linked to the visual appearance of the targets, our system promises broad impact on the efficiency and productivity of image analysts working in a variety of application areas.

ACKNOWLEDGMENT

The authors want to thank their colleagues from the University of California at Berkeley Visual Processing Laboratory, C. La, A. Garg, and J. Ales, for their help with data collection.

REFERENCES

[1] Y. Barniv, M. Aguilar, and E. Hasanbelliu, "Using EMG to anticipate head motion for virtual-environment applications," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 6, pp. 1087–1093, Jun. 2005.

[2] J. Beatty, "Task-evoked pupillary responses, processing load, and the structure of processing resources," *Psychol. Bull.*, vol. 91, no. 1, pp. 276–292, 1982.

[3] J. Beatty and L. W. Brennis, "The pupillary system," in *Handbook of Psychophysiology*, J. T. Cacioppo and L. G. Tassinary, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[4] R. F. Brem, "The never ending controversies of screening mammography," *Cancer*, vol. 3, no. 5, pp. 1549–1552, 2004.

[5] B. V. Dasarathy, *Decision Fusion*. Los Alamitos, CA: IEEE Comput. Soc. Press, 1994.

[6] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics," *J. Neurosci. Methods*, vol. 134, pp. 9–21, 2004.

[7] A. D. Gerson, L. C. Parra, and P. Sajda, "Cortical origins of response time variability during rapid discrimination of visual objects," *NeuroImage*, vol. 28, pp. 326–341, 2005.

[8] A. D. Gerson, L. C. Parra, and P. Sajda, "Cortical-coupled computer vision for rapid image search," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 174–179, Jun. 2006.

[9] A. K. Jain and D. Zongher, "Feature selection: Evaluation, application and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, Feb. 1997.

[10] W. Krenz, M. Robin, S. Barez, and L. W. Stark, "Neurology model of the normal and abnormal human pupil," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 10, pp. 817–825, Oct. 1985.

[11] I. Loewenfeld and O. Lowenstein, *The Pupil: Anatomy Physiology and Clinical Applications*. Detroit, MI: Wayne State Univ. Press, 1993.

[12] T. Marill and D. M. Green, "On the effectiveness of receptors in recognition system," *IEEE Trans. Inf. Theory*, vol. IT-9, no. 1, pp. 11–17, Jan. 1961.

[13] S. Mathan, D. Erdogmus, and Y. Huang, "Rapid image analysis using neural signals," presented at the Conf. Hum. Factors Comput. Syst., Montreal, QC, Canada, 2008.

[14] L. Parra, C. Spence, A. Gerson, and P. Sajda, "Response error correction—A demonstration of improved human–machine performance using real-time EEG monitoring," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 173–177, Jun. 2003.

[15] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, "Recipes for the linear analysis of EEG," *NeuroImage*, vol. 28, pp. 342–353, 2005.

[16] C. Privitera, L. Renninger, T. Carney, S. Klein, and M. Aguilar, "Pupil dilation during visual target detection," in *Proc. SPIE 20th Annu. Symp. Electron. Imag. Sci. Technol.*, Jan. 2008, pp. 68060T-1–68060T-11.

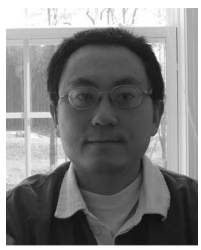
[17] P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, pp. 1119–1125, Nov. 1994.

[18] M. Qian, M. Aguilar, and L. Nolte, "The role of multi-level fusion in EEG-based classification for image triage tasks," to be published.

[19] A. R. Reibman and L. W. Nolte, "Design and performance comparison of distributed detection networks," *IEEE Tran. Aerosp. Electron. Syst.*, vol. 23, no. 6, pp. 789–797, Nov. 1987.

[20] J. D. Smith, G. A. Masek, L. Y. Ichinose, T. Watanabe, and L. W. Stark, "Single neuron activity in the pupillary system," *Brain Res.*, vol. 24, pp. 219–234, 1970.

[21] A. W. Whitney, "A direct method of nonparametric measurement selection," *IEEE Trans. Comput.*, vol. 20, no. 9, pp. 1100–1103, Sep. 1971.



Ming Qian received the B.A. degree in applied mathematics from Carroll College, Helena, MT, in 1994, the M.S. degree in electrical and computer engineering from Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, in 1997, and the M.S. degree in computer science and the Ph.D. degree in electrical and computer engineering from Duke University, Durham, NC, in 2004 and 2008, respectively.

From 1997 to 2001, he was a Design Engineer at Alcatel Network Systems, Raleigh, NC. From 2002 to 2003, he was a Research Associate in the Department of Neurobiology, and the Center for Neuroengineering, Duke University Medical Center, Durham. Since 2005, he has been a Research Scientist at Teledyne Scientific and Imaging LLC, Research Triangle Laboratory, Durham. His current research interests include biomedical signal processing, pattern recognition and classification, sensor fusion, and neural engineering.

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on December 3, 2009 at 19:30 from IEEE Xplore. Restrictions apply.

Mario Aguilar (M'93) received the B.S. degree in computer science from Jacksonville State University, Jacksonville, AL, in 1994, the M.A. and Ph.D. degrees in cognitive and neural systems from Boston University, Boston, MA, in 1993 and 1994, respectively.

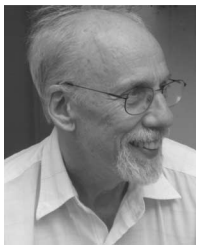
He was a Research Scientist with the Sensor Exploitation Group, Massachusetts Institute of Technology (MIT) Lincoln Laboratory. He was also an Associate Professor of computer science and psychology at Jacksonville State University. He is currently a Senior Scientist with the Intelligent Systems and Planning Group, Teledyne Scientific and Imaging LLC, Research Triangle Park Laboratory, Durham, NC. He is the author or coauthor of more than 40 scientific articles. His current research interests include sensor and information fusion, learning and pattern recognition, computer–human interaction, complex data visualization, and image analysis and understanding.

Karen N. Zachery received the M.S. degree in electrical engineering from the University of California, Davis, in 1997, and the M.S. degree in computer information systems from the University of Florida, Gainesville, in 1992.

She has more than 15-year experience in conducting R&D in the areas of digital signal/image processing, pattern recognition, and information/image exploitation. Her sensor expertise includes LADAR, IR, EO, MS/HS, SAR, GPSAR, and EMI. In the past eight years, her focus area has been multisensor fusion for target detection and recognition. She is currently with Teledyne Scientific and Imaging LLC, Research Triangle Laboratory, Durham, NC, where she is engaged in National Geospatial-Intelligence Agency's (NGA's) Biologically-Based Adaptive Multiband Feature Extraction and the Defense Advanced Research Projects Agency (DARPA's) Neurotechnology for Intelligent Analysts Programs. Her work experience also includes the development of a multisensor fusion performance model, landmine detection algorithms, and sensor fusion automatic target detection/recognition (ATD/R) algorithms.

Claudio Privitera received the Laurea degree in computer science in 1991 and graduated (*summa cum laude*) in 1995, both from the University of Pisa, Pisa, Italy.

From 1992 to 1995, he was with the Department of Informatics, Systems, and Telecommunications, University of Genoa, Italy, where he held a doctoral fellowship within the National Research Program on Bioelectronic Engineering. In 1994, he visited the Laboratorie Scribens, Ecole Polytechnique, Universite de Montreal. From 1995 to 1996, he was a Postdoctoral Researcher the International Computer Science Institute of Berkeley. In 1996, he joined the Neurology and Telerobotics Units, School of Optometry, University of California, Berkeley, as a Lecturer in the Mechanical Engineering Department. His current research interests include several aspects of biological and computational vision, eye movements, and visual attention. He is also engaged in the area of the human pupil system.



Stanley Klein received the B.S. degree in physics from California Institute of Technology (Caltech), Pasadena, in 1961, and the Ph.D. degree in particle physics from Brandeis University, Waltham, MA, in 1967.

Since 1987, he has been a Professor at the University of California, Berkeley. He switched research areas to visual neuroscience in 1973 with an emphasis on visual perception, psychophysical methodology, and nonlinear analysis and source identification using neural signals such as EEG, magnetoencephalogram (MEG), and functional MRI (fMRI).

Thom Carney received the Bachelor's degree in psychology from San Francisco State University, San Francisco, CA, in 1974, and the M.A. and Ph.D. degrees in physiological psychology from Stanford University, Palo Alto, CA, in 1983.

He was with the University of California, Berkeley, where he completed postdoctoral studies on binocular vision and motion perception and is currently involved in the study of visual perception using psychophysical methods and the brain imaging technologies: EEG, magnetoencephalogram (MEG), and functional MRI (fMRI).

Loren W. Nolte (S'56–M'57–LM'99–LSM'01) was born in Napoleon, OH. He received the B.S.E.E. degree from Northwestern University, Evanston, IL, and the M.S.E. and Ph.D. degrees in electrical engineering from The University of Michigan at Ann Arbor, Ann Arbor.

He was a Postdoctoral Researcher in signal detection theory at The University of Michigan at Ann Arbor for a year. He then joined the faculty at Duke University, Durham, NC, as an Assistant Professor of electrical engineering, and became a Professor of electrical and biomedical engineering in 1972 and was the Chairman of the Department of Electrical and Computer Engineering from 1998 to 2001. He has been a Visiting Professor at the University of Colorado, Boulder; Colorado State University, Fort Collins; the University of Washington, Seattle; and the Scripps Institution of Oceanography, La Jolla, CA. His research interests have included optimal Bayesian approaches to adaptive detection, classification, tracking, and localization of ocean acoustics signals, including recurrent transients. His recent contributions include the development of optimal physics-based array signal processing algorithms, which incorporate the physics, as well as the uncertainty of ocean acoustic environmental parameters. He was also an early contributor to optimal (likelihood-ratio-based) decision fusion. His current research interests include signal detection and estimation theory with applications to ocean acoustics, to biomedical statistical image processing in cancer research in collaboration with the Duke University Medical Center, and optimal decision fusion.

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on December 3, 2009 at 19:30 from IEEE Xplore. Restrictions apply.