# Rapid image analysis using neural signals

**Santosh Mathan**
Honeywell Laboratories
15001 NE 36th St
Redmond, WA 98052 USA
santosh.mathan@honeywell.com

**Deniz Erdogmus**
Oregon Health & Science Univ
20000 NW Walker Rd
Beaverton, OR 97006
derdogmus@ieee.org

**Catherine Huang**
Oregon Health & Science Univ
20000 NW Walker Rd
Beaverton, OR 97006
derdogmus@ieee.org

**Misha Pavel**
Oregon Health & Science Univ
20000 NW Walker Rd
Beaverton, OR 97006
derdogmus@ieee.org

**Patricia Ververs**
Honeywell Laboratories
7000 Columbia Gateway Dr
Columbia, MD 21046 USA
trish.ververs@honeywell.com

**James Carciofini**
Honeywell Laboratories
3660 Technology Dr
Minneapolis, MN 55418 USA
michael.dorniech@honeywell.com

**Michael Dorneich**
Honeywell Laboratories
3660 Technology Dr
Minneapolis, MN 55418 USA
michael.dorniech@honeywell.com

**Stephen Whitlow**
Honeywell Laboratories
3660 Technology Dr
Minneapolis, MN 55418 USA
michael.dorniech@honeywell.com


CHI 2008, April 5 – April 10, 2008, Florence, Italy
ACM 1-xxxxxxxxxxxxxxxxxx.

## Abstract
Extracting information from large collections of imagery is a difficult problem with few good solutions. Computers typically cannot interpret imagery as effectively as humans can, and manual analysis tools are slow. The research reported here explores the feasibility of speeding up manual image analysis by tapping into split second perceptual judgments using electroencephalograph sensors. Experimental results show that a combination of neurophysiological signals and overt physical responses — detected as a user views imagery in high speed bursts of approximately 10 images per second — provide a basis for detecting targets within large image sets. Results show an approximately six-fold, statistically significant, reduction in the time required to detect targets at high accuracy levels compared to conventional broad area image analysis.

## Keywords
Rapid Serial Visual Presentation (RSVP), EEG, Visual Search, Brain Computer Interface

## ACM Classification Keywords
H.5.2. Information interfaces and Presentation: User Interfaces: Interaction Styles H.3.3 Information Search and Retrieval: Information Filtering

## Introduction
The problem of searching for targets in vast collections of imagery affects practitioners in a variety of

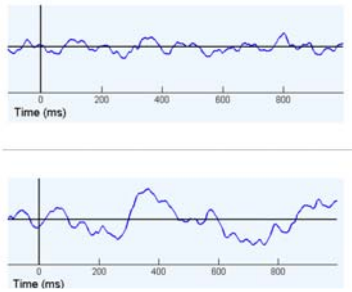


**Figure 1.** Baseline EEG (top) EEG segment containing an Evoked Response Potential (bottom)

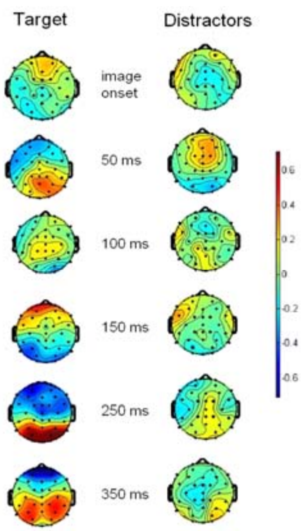


**Figure 2.** Average spatio - temporal pattern of electrical activity over the scalp following target (left) and distractor images (right)

domains— from medical diagnosis to intelligence image analysis. Advances in imaging and storage technology have lowered the cost of collecting and storing high volumes of imagery. However, the cost of searching through large sets of imagery for important information can be substantial. In many domains, effective search requires the expertise of highly skilled analysts. Unfortunately, the availability of skilled analysts is simply insufficient to cope with the volume of imagery to be analyzed.

The problems just highlighted have led to calls for triage techniques that can be used to rapidly screen high volumes of imagery. The triage process is a fast, preliminary examination of images to identify most targets, often with several false positives. Computer vision systems have been employed towards this end. However, in many contexts, these systems fall short of the sensitivity and specificity that humans display. Additionally, they cannot generalize to the extent that human analysts do.

Recently, researchers have begun exploring the feasibility of triage systems that leverage split-second human perceptual judgments. For example, Thorpe and colleagues asked participants to detect images of animals in a sequence of nature scenes presented for 20 milliseconds per image. This high-speed presentation modality is called Rapid Serial Visual Presentation (RSVP). Using EEG sensors, researches were able to detect a signal known as an event-related potential (ERP) within a few hundred milliseconds of the onset of target stimuli [6]. These findings point to the potential for using neurophysiological signals—specifically event-related potentials—as a way to detect targets within high speed sequences of images.

## Event Related Potentials

Event-related potentials refer to a morphological change in EEG waveforms in response to rare task-relevant stimuli. They are typically measured by inspecting EEG activity within a window of several hundred milliseconds following critical events. Figure 1 shows EEG activity at a particular sensor following a task-irrelevant stimulus (distractor) and a task-relevant stimulus (target). The x-axis depicts the progression of time, in milliseconds, following the stimulus (the zero point corresponds to the onset of a stimulus). The wave form associated with the target shows a pronounced amplitude perturbation within a few hundred milliseconds of stimulus onset.

Research suggests that ERPs reflect the activity of cognitive processes necessary for processing and coordinating a response to task-relevant stimuli. The brain's response to critical events, such as the presence of targets, begin in frontal areas—generating top-down, intent information—and propagate to sensorimotor areas—triggering events that regulate bottom up information transmission through sensory and response selection areas [4].

ERPs are difficult to detect. These signals typically range in amplitude from approximately 1 to 10 microvolts, while background EEG activity may range from 10 to 100 microvolts. Common events such as eye blinks or facial muscle activity can completely obscure ERPs. In order to deal with such a low signal-to-noise ratio, ERP detection has relied on a strategy of trial averaging. Under this strategy, an experimental stimulus is presented to a participant several times. The waveforms elicited by each stimulus are averaged. Background EEG washes out in the averaging process

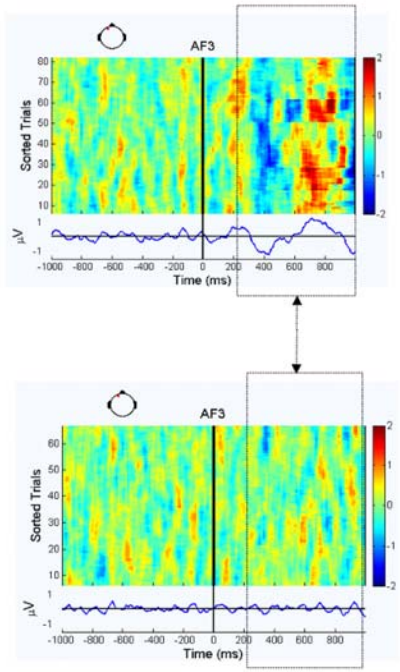


**Figure 3.** Electrical activity following target (top) and distractor images (bottom) at a single EEG site.

Y-axis in each plot indexes individual trials or epochs; the x-axis represents time. The 0 point on the x-axis (bold, black line) represents time of stimulus onset. Color represents the polarity and amplitude of electrical activity.

There is a consistent pattern of sustained, high amplitude activity following stimulus onset at practically every trial containing a target.

and the event-induced activity becomes prominent. While integrating information across repeated presentations of a stimulus is an effective way to identify ERPs, it is an impractical strategy for a triage platform. Repeated presentation of stimuli compromises the efficiency of the search process.

Recently, researchers have developed promising approaches for single-trial ERP detection. Instead of integrating sensor data over time, they rely on integrating information spatially, across EEG sensors. Spatial integration of EEG data around a window of a few hundred milliseconds following an image trigger can provide a basis for accurate, single-trial ERP detection [3,5]. Both linear and nonlinear classification approaches are effective in detecting ERPs based on spatio-temporal activation patterns across sensors.

## Neurotechnology for Image Analysis

Recent research has explored the feasibility of using ERPs to detect targets within high speed presentation of images. These studies show promising results. For example, in a recent study, the authors of this paper asked participants to detect boats and ships within a sequence of images extracted from a broad area satellite image of a peninsula [5]. Qualitative analysis of the EEG data revealed a clear pattern of spatio-temporal activity that could serve to discriminate between images containing targets and distractors—starting at approximately 150 ms following stimulus onset (Figure 2). The analysis also revealed that trial-to-trial variability of EEG samples associated with each image class (target vs. distractor) was low relative to the variability between classes (Figure 3).

The study just described also examined the feasibility of accurate, single-trial detection of ERPs in the context of complex satellite imagery. The study included three twenty-minute sessions of image analysis spanning the course of an hour. A support vector machine classifier that was trained on data from the first twenty minute session was able to classify samples from the third twenty minute session with a very high degree of accuracy (area under the receiver operator characteristic curve of 0.90 or higher). Practically, this is an important finding. Analysts anecdotally report analyzing imagery for spans of approximately an hour. However, prior work focused largely on data collected over the span of sessions separated by under 10 minutes. This study also demonstrated that reliable, single-trial ERP-based target detection was possible with relatively practical 32 electrode EEG systems. In contrast, much of the prior work in this area used arrays of 64 or more electrodes.

*Relevance of overt responses*

While the preceding discussion has focused on neural signals, overt physical responses such as key presses can also provide accurate triage performance in an RSVP context, but the latency and variability associated with motor responses are higher than with ERPs. Consequently, ERPs can provide a better basis for precise localization of targets within high-speed image sequences. Despite lower temporal resolution, motor responses could still provide a redundant source of information to point to broad regions of high target likelihood. Important individual differences must also be considered; researchers have observed that physical responses can provide a better basis for target detection for some individuals [3]. Considering these factors, the triage system discussed here relies on a

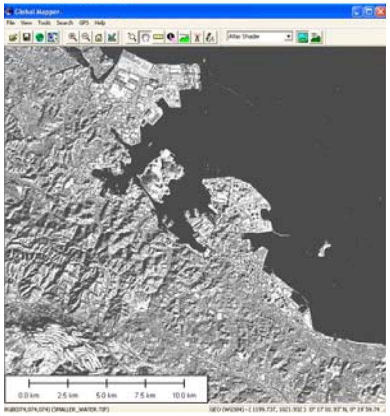

**Figure 4:** Global Mapper, a broad area image analysis tool


**Figure 5:** Image presentation using RSVP presentation modality. Images were presented at a rate of 100 ms / image in short bursts lasting 5 seconds.

fusion of ERPs and overt physical responses—with ERPs weighted substantially higher to exploit the higher degree of temporal resolution.

While studies point to the feasibility of using neurophysiological signals as the basis for image triage, the efficiency of neurophysiologically-driven image triage compared to conventional broad area image analysis tools is generally unknown. The research reported below compares the efficiency gains associated with neurophysiological image triage to target detection using conventional image analysis tools among a group of professional image analysts.

## Experimental Method

*Participants*

Seven military image analysts (IA) participated in an experiment comparing neurophysiologically driven image triage to conventional broad area image search. All participants had experience with a broad range of imagery and target types. They were all trained in the use of geo-spatial analysis tools. None of the participants were familiar with the RSVP modality.

*Task*

We employed a within subjects experimental design to compare broad area search to RSVP triage (with counterbalancing of each task condition). In each condition, participants searched for surface to air missile sites within grayscale commercial, satellite imagery (Figure 6). The average image represented an area of over 200 sq km. These large images, tens of thousands of pixels wide and tall, were decomposed into image chips or tiles that were 400 x 400 pixels in width and height. Each chip represented an area of 0.09 sq km.

*Baseline Condition*

In the baseline condition, participants used a geospatial analysis tool called GlobalMapper (Global Mapper Software LLC, Olathe, Kansas), that allows high resolution satellite imagery to be efficiently searched and annotated. GlobalMapper provides zoom and pan controls to search large, high-resolution images (Figure 4). Participants were allowed as much time as they wished to search each image for targets. All participants were shown prototype images depicting the targets during the

*RSVP Condition*

In the RSVP condition, each broad area image was decomposed into chips presented to participants at a rate of 100 milliseconds per chip. One participant processed images at a rate of 150 milliseconds per chip because of poor performance during training at the 100 millisecond rate. Images were presented in short bursts approximately 5 seconds duration (Figure 4, right). Participants were asked to press a key as soon as a target was seen. To break monotony and minimize possible eye strain, consecutive trial blocks were separated by a fixation screen of user-controlled duration.

Chips were presented on a 21 inch, CRT monitor with a screen resolution of 1240 x 768 pixels. Participants could position themselves at a comfortable distance from the screen. All images shared a similar level of luminance and were presented using a script developed for Presentation®, a stimulus presentation tool developed by Neurobehavioral Systems, Albany, CA.

The RSVP sessions were structured in two phases: a training phase, designed to familiarize participants with

**Figure 6:** Examples of surface to air missile sites (SAM) -- targets sought by participants. SAM sites typically consist of a few missile launchers in a radial arrangement, with paths linking them.

detecting targets under different RSVP rates, and a performance phase. In the training phase, participants viewed images at a rate of 10 images per second in five-second bursts. These blocks contained non-target images drawn randomly from several broad area images. Target examples were randomly inserted into half the trial blocks. Participants responded with a key press as soon as a target was detected. Participants received feedback on their responses at the end of each trial block. In performance mode, the chips were presented in the natural spatial order in which they occur in the broad area image. Participants received no feedback in performance mode.

*EEG Acquisition*

In the RSVP condition EEG was collected using a BioSemi® ActiveTwo® amplifier with 32 electrodes (BioSemi, Amsterdam, Netherlands). Channels were sampled at 256 Hz. Triggers sent by the Presentation script to mark the onset of target and distractor stimuli were received by the BioSemi system over a parallel port and recorded concurrently with EEG signals. User key presses, indicating the presence of targets, were also recorded using the BioSemi system. EEG was bandpass filtered between 1 Hz and 30 Hz using an 8th order Butterworth digital filter.

*EEG Segmentation and Classification*

As mentioned earlier, a trigger or brief pulse was sent to the EEG amplifier with each image that was displayed to the participant. The EEG amplifier also recorded pulses associated with key presses. A segment of EEG data and key press data was extracted around each image trigger. These segments, referred to as epochs, contained a second of EEG and key press data on either side of each image trigger. EEG and key press epochs associated with target images and non-target images were extracted from training phase data.

Epochs extracted from training phase data were used for classifier training. A support vector machine (SVM) classifier [2] trained on training phase data was used to classify epochs associated with each image in performance mode. Support vector machines are a widely-used linear machine-learning technique that relies on ideas from statistical learning theory to provide good generalization performance. Support vector machines can also be used in the context of problems that are not linearly separable by projecting data into a higher dimensional space where the data may be linearly separable. This study used a non-linear support vector machine with a radial basis function kernel.

## Results

For each participant, two support vector machine classifiers were trained with training phase RSVP data. One classifier was based on EEG epochs; the other was based on key press epochs. These classifiers were used to classify performance phase images as either targets or distractors. Outputs from the EEG and key press classifiers were re-scaled to lie between 0 and 1 and fused using a weighted combination of values. Because of the higher temporal resolution associated with EEG, outputs of the EEG classifier were weighted twice as high as the key press classifier. The fused values were rescaled to lie between 0 and 1 and interpreted as approximate indicators of the probability of a given image being a target.

The probability estimates from the classifiers were used to generate probability maps. These contour maps were

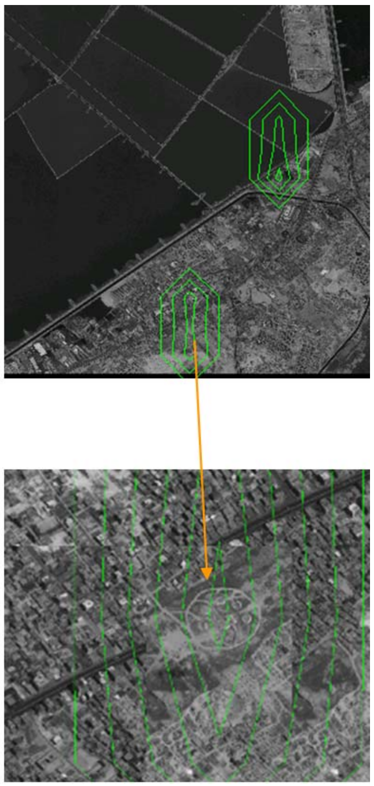

**Figure 7:** High resolution satellite images are annotated with probability contours depicting target hotspots. Analysts can confirm the presence of targets by zooming in.

overlaid on a broad area image; users could identify targets by zooming into probabilistic hotspots (Figure 7). Hotspots overlapping targets are considered hits; targets without an overlapping hotspot are regarded as misses.

Analysis of analyst performance shows close to a 6 fold, statistically significant ($F(1,12) = 239.09$, $p < 0.001$) reduction in image analysis time relative to baseline search. Participants were able to scan images at an average rate of .86 sq km/sec ($\sigma = 0.11$) in the triage condition, compared to 0.15 sq km/sec ($\sigma = 0.04$) in the baseline condition. This increase in throughput is achieved without compromising target detection accuracy. On average, participants in the triage condition detected 96% ($\sigma = 7\%$) of targets, compared to 87% ($\sigma = 17\%$) in the baseline condition. The false positive rate in both conditions was low (average of 3.9 false positives in the triage condition and 0.8 in the baseline). Neither the detection rate, nor the false positive rate were statistically significant

While the efficiency gains presented here are quite large, there are several issues that have to be addressed in future work. First, detection accuracy has the potential to vary as targets move away from the user's fixation point. The presentation durations described here, simply do not afford enough time for eye saccades to search within each chip. Practical implementations of RSVP-based triage systems should consider chips with overlapping content or employ intelligent image segmentation and orienting algorithms as a pre-processing step. Other aspects of the imagery such as the scale of each chip relative to the target, and the degree of visual clutter in a scene can also compromise performance. Second, a variety of physical and cognitive states –- ranging from low attention levels and high working memory load, to eye blinks and gross head movements can affect performance. We are currently exploring the efficacy of EEG based monitoring of attention and working memory levels – desktop eye trackers are being evaluated to detect head and eye activity.

## Acknowledgements

This research was funded by the Defense Advanced Research Projects Agency (DARPA) under contract HM1582-05-C-0046.

## References

[1] Delorme A and Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics J Neuroscience Methods, 134:9-21, 2004

[2] Duda, R.O, Hart, R. E. and Stork D. G. Pattern Classification, Second Edition. John Wiley & Sons, New York, 2001

[3] Gerson, A.D., Parra, L.C., and Sajda, P. (2005) Cortical Origins of Response Time Variability during Rapid Discrimination of Visual Objects, NeuroImage, 28 (2) 326-341.

[4] Makeig S, Westerfield M, Jung T-P, Enghoff S, Townsend J (2002) Dynamic brain sources of visual evoked responses. Science 295: 690–693

[5] Mathan, S., Whitlow, S., Erdogmus, D., Pavel, M., Ververs, P., Dorneich, M. (2006) Neurophysiologically driven image triage: a pilot study. Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems 2006 v.2 p.1085-1090.

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