Neurotechnology for Image Analysis: Searching For Needles in Haystacks Efficiently

Santosh Mathan, Patricia Ververs, Michael Dorneich, Stephen Whitlow, Jim Carciofini

Honeywell Laboratories
3660 Technology Dr, Minneapolis, MN 55418
santosh.mathan@honeywell.com

Deniz Erdogmus, Misha Pavel, Catherine Huang, Tian Lan, Andre Adami

Oregon Health and Science University
20000 NW Walker Rd, Beaverton, OR 97124
derdogmus@ieee.org

Abstract

Timely and efficient processing of complex imagery is a vital aspect of important domains such as intelligence image analysis. As technological developments lower the cost of gathering and storing imagery, the cost of searching through large image sets for important information has been growing substantially. This paper demonstrates the feasibility of using neurophysiological signals associated with early perceptual processing to identify critical information within large image sets efficiently. Experimental results show that a combination of neurophysiological signals called evoked response potentials and overt physical responses, detected in conjunction with high speed presentation of images, provide a basis for detecting targets within large image sets efficiently and accurately. Experimental evaluations of neurophysiologically driven image triage show over a five-fold, statistically significant, reduction in the time required to detect targets at high accuracy levels compared to conventional broad area image analysis.

1 INTRODUCTION

The problem of searching for targets in vast collections of imagery affects practitioners in a variety of 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 often be substantial. In many domains, such as intelligence analysis, effective searches currently require 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 (Kenyon, 2003). As a result, most intelligence imagery remains unexamined.

The problems just highlighted have led to calls for effective triage techniques that can be used to rapidly screen high volumes of imagery and identify a subset of images that merit careful scrutiny by an image analyst (Kenyon, 2003). A triage process trades off specificity in favor of sensitivity. 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 (e.g. Collins 2000). However, in many contexts, these systems fall short of the sensitivity and specificity that humans display, and they cannot generalize to the extent that human analysts do.

Recently, researchers have begun exploring the feasibility of triage systems that leverage human visual processing capabilities while raising the efficiency associated with the manual search process. One promising avenue for realizing an efficient triage system may lie in electroencephalogram (EEG) signals recorded in conjunction with rapid serial visual presentation (RSVP) of images. For example, Thorpe and colleagues asked participants to detect images of animals in a sequence of nature scenes presented for 20 milliseconds per image. Using EEG sensors, researches were able to detect a brain signal known as an evoked response potential (ERP) within 150 ms of the onset of target stimuli (Thorpe, Fize & Marlot, 1996). These findings point to the potential for using neurophysiological signals—specifically evoked response potentials—as a way to detect targets within high speed sequences of images.

1.1 Evoked Response Potentials

Evoked response potentials refer to a morphological change in EEG waveforms in response to 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 following the stimulus in milliseconds (the zero point corresponds to the onset of a stimulus). The wave form associated with the target shows a pronounced amplitude perturbation following stimulus onset.

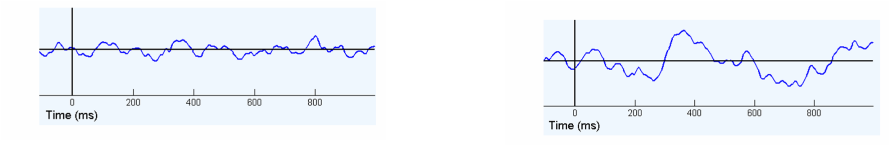


Figure 1. Baseline EEG (left) EEG segment containing an evoked response potential (right)

Research suggests that ERPs reflect the activity of underlying 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, may 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 (Makeig, Westerfield, Jung, Enghoff, & Townsend, 2002).

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, 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 application domains, such as a triage platform. Repeated presentation of stimuli compromises the efficiency of the search process. In domains where efficient ERP detection is critical, accurate detection of ERPs within a single trial becomes necessary. Recently, researchers have developed promising approaches for single-trial ERP detection (e.g. Parra et al., 2003 and Gerson, Parra, & Sajda, 2005). 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 (Mathan et al, 2006). Both linear and nonlinear classification approaches are effective in detecting ERPs based on spatio-temporal activation patterns across sensors.

1.2 Neurophysiologically Triggered Image Triage

Recently, researchers have begun exploring 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 (Mathan et al, 2006). The image was provided by the National Geospatial-Intelligence Agency. Qualitative analysis of the neurophysiological data revealed a clear pattern of spatio-temporal EEG activity starting approximately 150 ms following stimulus onset that could discriminate between images containing targets from distractors (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).

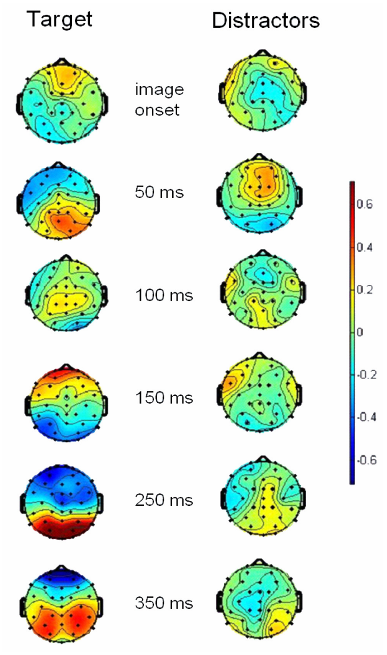


Figure 2. Average spatio-temporal pattern of electrical activity over the scalp following target (left) and distractor images (right).

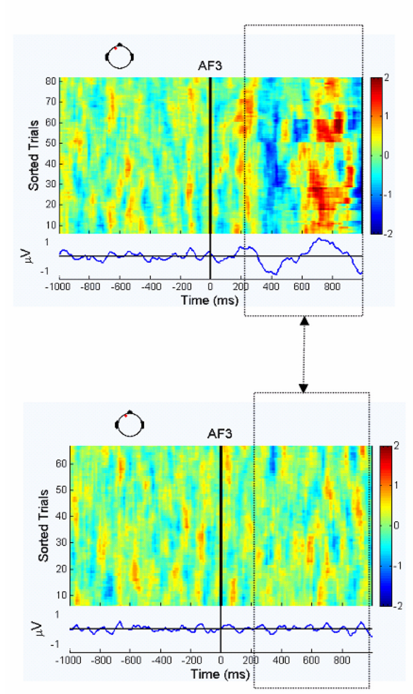


Figure 3. Activity following targets (top) and distractors (bottom) at a single EEG site. Individual epochs are stacked to show mean activity and variability.

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. Prior work focused largely on data collected over the span of sessions separated by under 10 minutes. However, analysts anecdotally report analyzing imagery for spans of approximately an hour. This study also demonstrated that reliable single-trial ERP-based target detection was possible with relatively practical 32 electrode EEG systems; much of the prior work in this area used arrays of 64 or more electrodes.

1.3 Relevance of Overt Physical Responses

Note that 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 (Gerson, Parra, & Sajda, 2006). Considering these factors, the triage system discussed here relies on a fusion of ERPs and overt physical responses, with ERPs weighted substantially higher to exploit the higher degree of temporal resolution. Details of the fusion approach are discussed below.

1.4 Research Objective: Comparing Search Modalities

While studies point to the feasibility of using neurophysiological signals associated with perceptual judgments for image triage, the relative 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.

2 METHOD

This paper focuses on an experiment comparing the efficiency of searching for targets within broad area satellite images using two techniques: broad area search using geospatial image analysis tools, and search using a combination of neurophysiological signals and overt physical responses in the context of RSVP. The experimental evaluation employed a single factor (broad area vs. RSVP search), between-subjects experimental design.

2.1 Imagery

Participants in both the broad area and RSVP search conditions searched for three types of targets in a broad area satellite image of a peninsula. Imagery was provided by the National Geospatial Intelligence Agency (NGA). For the RSVP condition, the broad area image was decomposed into 783 image chips. These chips were also provided by the NGA. Other than rescaling the chips so that each chip could be viewed at a glance, the chipped images were not digitally manipulated in any way. Participants were asked to detect three types of targets at three different scales: three golf courses spanning four or more chips, an oil storage depot that spanned two chips, and an oil tanker that was fully contained within a chip.

2.2 Baseline

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, left). Participants were given as much time as they wished to become familiar with the tool using a broad area satellite image that was not the focus of this experiment. All participants were shown prototype images depicting the targets and told exactly how many instances of each target were present in the broad area image. The prototype images consisted of the targets participants were looking for with some of the surrounding geographical contextual features removed.

Seventeen participants, recruited from the engineers and scientists at Honeywell Laboratories, participated in the experiment. Participants were asked to try to detect targets as quickly as possible. However, no time limits were placed on the search. Participants could detect targets in any order they wished, and the time elapsed from the beginning of the experiment was logged with each target detected.

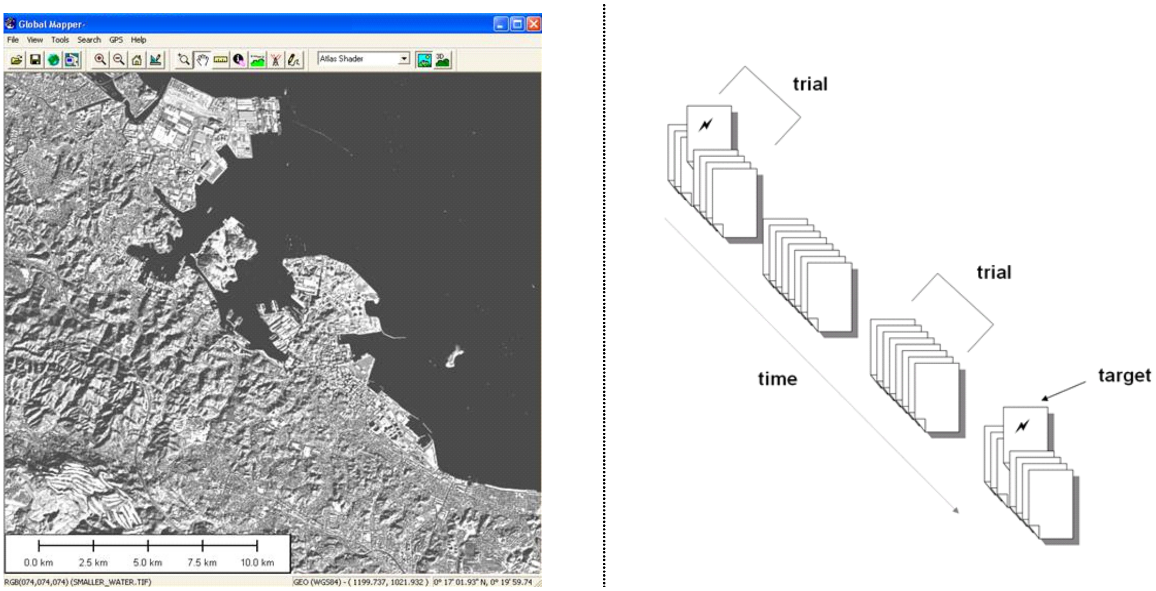

Global Mapper
File View Tools Search GPS Help
Atlas Shader
0.0 km
2.5 km
5.0 km
7.5 km
10.0 km
RGB(074,074,074) (SMALLER_WATER.TIF)
GEO (WGS84) - (1199.737, 1021.932)
0° 17' 01.93" N, 0° 19' 59.74
trial
trial
time
target

Figure 4. Global Mapper, a broad area image analysis tool (left). Image presentation using RSVP modality (right).

2.3 RSVP search

In the RSVP condition, the image chips described above were presented to participants at rates rates of: 75, 100, 150, and 200 milliseconds per image. Images were presented in short bursts or trial blocks of approximately 5 seconds duration (Figure 4, right). Participants were asked to press a key to indicate the presence of a target. To break monotony and minimize possible eye strain, consecutive trial blocks were separated by a fixation screen of user-controlled duration.

2.3.1 Display

Images of 400 x 400 pixels 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.

2.3.2 Data Acquisition

EEG was collected using a BioSemi® ActiveTwo® amplifier (BioSemi, Amsterdam, Netherlands) using 32 electrodes. 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.

2.3.3 Participants and Session Structure

Six participants, graduate students from Oregon Health and Science University, participated in the study. The RSVP experimental sessions were structured in two phases: a training phase, designed to familiarize participants with detecting targets under different RSVP rates, and a performance phase. In the training phase, participants viewed images in five-second trial blocks. These blocks contained non-target images drawn randomly from the peninsula chip set. One of the target prototypes was 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. RSVP rates in training and performance modes included 75ms, 100ms, 150ms, and 200ms per image.

2.3.4 Data 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 the training phase data.

Epochs extracted from training phase data were used for classifier training. A support vector machine (SVM) [2] classifier 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.

3 RESULTS

3.1 Baseline: Broad Area Image Analysis

All 17 participants in the baseline condition were able to detect each of the five targets in the broad area satellite image. On average participants took 11.14 min to detect all targets, (SD = 6.24 min). Participant performance ranged from a minimum of 3.68 minutes to a maximum of 24.1 minutes, with a median detection time of 11.18 min. It is important to note that these times may underestimate the time required to process the broad area image in realistic search contexts. Participants knew exactly how many targets were in the image and could terminate the search as soon as all targets were found. In the absence of knowledge about the precise number of targets—which is typical in many application contexts—it may take subjects considerably longer to terminate the search.

Target type made a difference in the time taken to detect a target. Golf courses that are clearly visible without magnification were detected most easily by most participants. The oil tanker and oil storage depot required a systematic search of the image with magnification and panning, and participants took a significantly longer time to detect them.

3.2 RSVP Search

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. Outputs from the two classifiers were fused using a weighted combination of output 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.

Probability values that lie in the immediate vicinity of key presses were rendered on a contour map. Contour clusters served to indicate the most likely location of targets. These contour maps could be overlaid on a broad area image; users could identify targets by zooming into high probability regions. Figure 5 depicts a probability map for one participant: white squares denote location of targets; colored contour areas depict likely location of targets estimated by the classifiers.

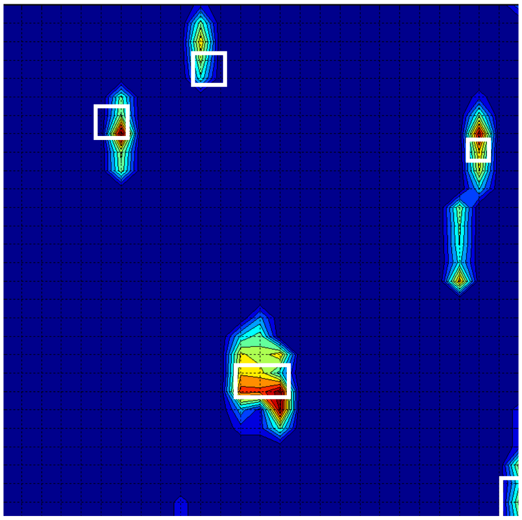


Figure 5. Contour map displaying regions likely to contain targets. White boxes (displayed here for the reader and not shown to participants) depict locations of targets in the image. These contours could be overlaid on the broad area satellite image and point to areas that the analyst should scrutinize closely.

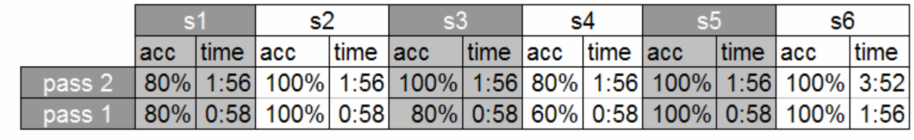
3.2.1 RSVP Search Performance

Each participant scanned the broad area image twice in performance mode at each RSVP rate (75ms, 100ms, 150ms and 200ms per image). The analysis reported here focuses on performance associated with the fastest performance rate for each participant. The fastest available rate for 5 out of 6 participants was 75ms per image (0:58 minutes to process 783 chips in a systematic sweep of the broad area image). The fastest available rate for one participant was 150ms because of data lost due to logging errors (1:56 minutes to process 783 chips).

High probability clusters that overlap target locations were counted as detections in this analysis. False alarm rates were low for most participants: ranging from 0 to a maximum of 4 clusters. The median false positive cluster rate was 0.

Table 1 depicts the peak accuracy level reached after each pass or sweep for each participant—the table also depicts the time elapsed for each pass. The table indicates that all participants cross the 80% accuracy level within two passes. Four out of six participants cross the 100% detection threshold within two passes.

Table 1. RSVP mode performance for each RSVP participant. Table depicts detection accuracy and elapsed time associated with each pass.



	s1		s2		s3		s4		s5		s6	
	acc	time	acc	time	acc	time	acc	time	acc	time	acc	time
pass 2	80%	1:56	100%	1:56	100%	1:56	80%	1:56	100%	1:56	100%	3:52
pass 1	80%	0:58	100%	0:58	80%	0:58	60%	0:58	100%	0:58	100%	1:56

.

As Figure 6 (left) depicts, participants in the baseline condition took 7.20 minutes on average (SD = 4.55 minutes) to reach the 80% target detection level, compared to 1.33 minutes in the RSVP condition (SD = 0.51 minutes). These differences were statistically significant: F (1, 21) = 9.63; p = .0053. Four out of six participants in the RSVP condition reached the 100% detection level within two passes. Figure 6 (right) shows that RSVP participants who reached the 100% detection level within two passes took 1.5 minutes on average (SD = 0.57 minutes), compared to an average of 11.6 in the baseline condition (SD = 6.25 minutes) [F(1,19) = 10.10; p = .0049].

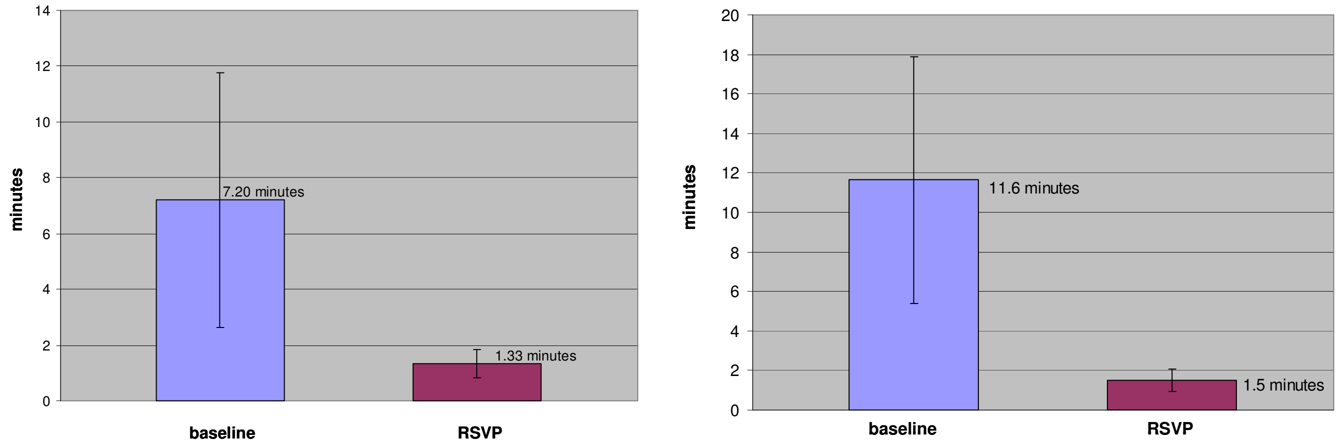


Figure 6. Comparison of time required to reach 80% detection level (left) and 100% detection level (right) in baseline and RSVP conditions. Error bars denote standard deviation.

4 DISCUSSION

The results presented above show a substantial efficiency gain associated with the RSVP condition. The analysis reveals a five-fold reduction in the time required for target detection at the 80% level in the RSVP condition compared to the baseline. The analysis also reveals a seven-fold reduction in the time required to reach the 100% detection level in the RSVP condition. These findings demonstrate the viability of neurophysiologically triggered image triage as a human computer interaction modality for efficiently searching through high volumes of imagery. Processing images at the rates reported here would allow tens of thousands of images to be screened with a high degree of accuracy within the span of an hour.

As mentioned earlier, the time required to identify targets in the baseline condition are likely to be an underestimate. Subjects were told exactly how many targets were present in the image and could terminate the search as soon as all targets had been found. Search termination could have taken considerably longer without knowledge of the number of targets present.

While the efficiency gains presented here are quite large, two out of the six participants could not exceed the 80% detection level following two passes in the RSVP condition. There are several factors that could have contributed to this outcome:

- First, participants had difficulty detecting targets that were offset from the center of the screen. Two types of targets posed the most difficulty in the RSVP condition: a golf course and the oil tanker. The problematic golf course lay in the boundary between four chips. Visual features of the course occupied the periphery of each chip, making it harder to detect if a user was fixating on the center of the screen. Similarly, the oil tanker was offset from the center of the screen and occupied a relatively small portion of the image. 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.

- Second, a variety of user states can affect the ability of a user to detect targets within high speed sequences of images. It is natural for attention levels to wax and wane over the course of an analysis session. Events such as eye blinks that occur several times a minute and last several hundred milliseconds can prevent several images from being processed appropriately. Inappropriate gaze and head orientations can also compromise effective processing. User state monitoring algorithms that detect sub-optimal user states could play a mitigating role. For example, presentation rates of images could be varied to match the processing capacity of users. Additionally, images that are judged to be inappropriately processed could be flagged for follow-on review.

- Third, familiarity with the RSVP presentation modality may have implications for triage performance. Our participants were naïve to processing high speed image sequences. It is conceivable that performance could improve as a function of training and experience. It is also possible that individual differences in the ability to detect targets within high speed image sequences could preclude some users from being effective in RSVP tasks.

Future efforts will seek to extend the work reported here by evaluating the efficiency of RSVP based triage with trained image analysts. Our focus will also shift to developing and evaluating software algorithms that monitor the state of the user and adapt the system to mitigate the impact of sub-optimal states.

5 ACKNOWLEDGEMENTS

This research was funded by the Defense Advanced Research Projects Agency (DARPA) and the National Geospatial Intelligence Agency (NGA). The points expressed here are those of the authors and do not necessarily reflect the views of DARPA or NGA.

6 REFERENCES

Collins, Lipton, Kanade, Fujiyoshi, Duggins, Tsin, Tolliver, Enomoto, & Hasegawa. (2000). A System for video surveillance and monitoring: VSAM final report, Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May, 2000

Delorme, A. & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics, *J Neuroscience Methods*, 134:9-21

Duda, R.O, Hart, R. E. and Stork, D. G. (2001). Pattern classification (2d Ed.). John Wiley & Sons, New York, NY.

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

Gerson, A.D., Parra, L.C., & Sajda, P. (2006) Cortically-coupled Computer Vision for Rapid Image Search", *IEEE Transactions on Neural Systems & Rehabilitation Engineering*. 14(2). 174-179

Kenyon, H. S. (2003). Unconventional information operations shorten wars. *Signal Magazine*. Armed Forces Communications and Electronics Association. 57:33-36

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

Mathan, S., Whitlow, S., Erdogmus, D., Pavel, M., Ververs, P., Dorneich, M. (2006) Neurophysiologically driven image triage. CHI '06 extended abstracts on human factors in computing systems. *Proceedings of the 2006 Conference on Human Factors in Computing Systems*. Montreal, Canada. 1085 – 1090.

Parra, L., Alvino, C., Tang, A., Pearlmutter, B., Yeung, N., Osman, A., & Sajda, P. (2003), Single trial detection in EEG and MEG: keeping it linear. *Neurocomputing*. (52-54) 177-183.

Spence, R., Witkowski, M., Craft, B., de Bruijn, O. & Fawcett, C. (2004) Image Presentation in Space and Time: Errors, Preferences and Eye-gaze Activity. International Conference on Advanced Visual Interfaces (AVI-2004). pp. 141-149.

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