Characterizing the Performance Limits of High Speed Image Triage Using Bayesian Search Theory

Santosh Mathan[1], Kenneth Hild[2], Yonghong Huang[2], and Misha Pavel[2]

[1] Honeywell Laboratories, 15001 NE 36th St., Redmond, WA 98052, USA
[2] Oregon Health and Science University, 3303 SW Bond Ave, Portland, OR 97239, USA
santosh.mathan@honeywell.com, k.hild@ieee.org,
huang@csee.ogi.edu, pavel@bme.ogi.edu

Abstract. The rapid serial visual presentation (RSVP) modality has been used in conjunction with neurophysiological and behavioral responses to identify targets within large volumes of imagery efficiently. The research reported here uses optimal search theory to characterize the limits of this approach. Search theory is used to inform the estimation of detection functions. These functions provide a principled basis for selecting presentation parameters that balance search efficiency and accuracy. Detection functions are also used to characterize individual differences in search performance and to assess the extent to which the RSVP presentation modality generalizes across a class of complex targets.

Keywords: EEG, Search Theory, Rapid Serial Visual Presentation, Visual Psychophysics, Detection Functions, Target Detection.

1 Introduction

The challenge of finding information in large volumes of imagery has few good solutions. Both automated and manual approaches to target detection have limitations. Computer vision-based solutions often can't deal with novelty or variability, nor can they exploit contextual information and prior knowledge to the extent that humans can. On the other hand, manual image analysis is slow—requiring careful, methodical scrutiny of an image to identify potential targets. Manual analysis tools typically engage slow and deliberate, top-down cognitive processes that operate on a time scale of seconds or tens of seconds. These limitations of conventional search techniques have practical implications in a wide variety of domains; from military image analysis, to geospatial and medical imagery analysis. The volume of imagery available in these domains far exceeds the resources available to process them.

1.1 Tapping into Split-Second Perceptual Judgments

One avenue for raising the efficiency of the manual search process is to exploit the fast, automatic, bottom-up perceptual judgments that people make routinely. Specific examples include the perceptual processes engaged in returning a tennis serve, hitting a baseball, or reacting to an obstacle on the highway while driving—we can detect critical events and initiate physical responses to them in a couple of hundred

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 95–103, 2011.

milliseconds. Yet, these processes can solve complex perceptual recognition problems, including those that appear to require cognitive interpretation [1]. In recent years, researchers have attempted to take advantage of these processes to boost the efficiency of manual search. Research has shown that a combination of the rapid serial visual presentation (RSVP) technique and the event-related potential (ERP) signal detected using electroencephalograph (EEG) sensors can provide a way to identify targets in imagery using split-second perceptual judgments [2]. RSVP is a presentation paradigm under which a sequence of images is flashed to users at very high rates—where each image is presented for durations spanning just a few tens or hundreds of milliseconds (Figure 1).

Our research has shown that the RSVP presentation modality, when employed in the context of a multi-stage search process, can help professional image analysts identify a broad range of complex targets in high-resolution satellite imagery efficiently and accurately [4]. In the first stage of our approach, broad area images, spanning tens of thousands of pixels in width and height are decomposed into chips a few hundred pixels wide and tall. Each image chip must be scaled appropriately to the dimensions of the target of interest, as the high presentation rates used in the RSVP paradigm preclude eye saccades to search an image. In the first stage of the search process, these chips are presented to users in high speed bursts—anywhere from 2 to 15 chips per second. EEG sensors record neural responses to each chip. Images that elicit an ERP signal are classified as potential targets. In the second stage of the search process, users examine the subset of images identified as being targets in the RSVP-based search and eliminate false positives.

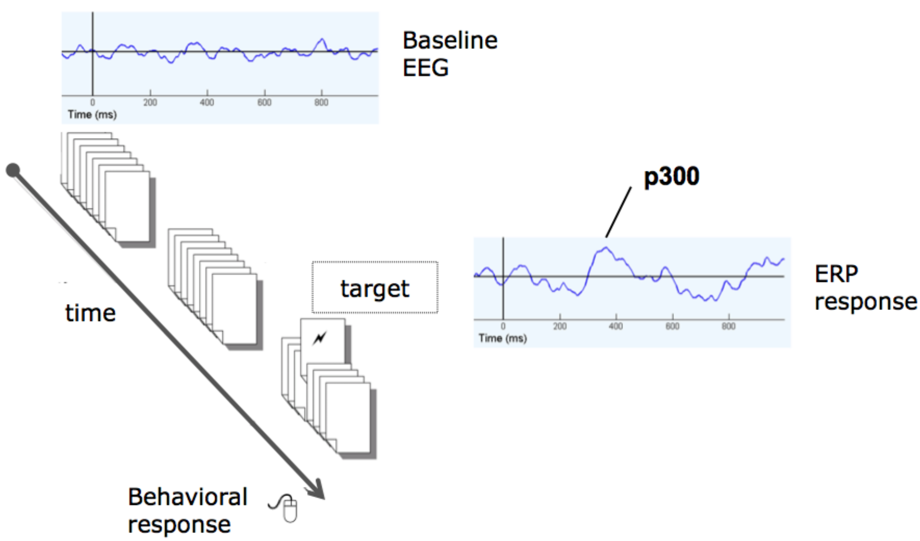

Fig. 1. Large high-resolution images are decomposed into chips and presented to users in high-speed bursts lasting 2 to 5 seconds. Patterns of EEG and other responses are used to detect likely targets in the image sequence.

Empirical evaluations of the approach outlined above show that the RSVP presentation modality in conjunction with response signals ranging from EEG [2, 4, 5, 7] to button clicks [7, 4] and pupil dilation [6] can provide an efficient and accurate basis for identifying a broad range of targets in natural imagery. For instance, in an experiment involving experienced military image analysts, an RSVP-based search for surface-to-air missile sites contributed to a six-fold reduction in processing time relative to conventional search [4]. Participants were able to scan images at an average rate of .86 sq km/sec (s.d = 0.11) in the RSVP condition, compared to 0.15 sq km/sec (s.d = 0.04) in the conventional search condition. This increase in processing efficiency was achieved without compromising target detection accuracy. On average, participants in the RSVP condition detected 96% (s.d = 7%) of targets, compared to 87% (s.d = 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 differences in detection rate, nor the false positive rate were statistically significant.

While demonstrations of the efficacy of an RSVP-based triage approach are promising, important questions about the generality of the observed results must be addressed in order for this technique to be broadly adopted. For example, it is important to determine whether the observed efficiency gains extend beyond the specific targets used in a given experiment to other targets of a particular class. Additionally, it is essential to determine whether the observed results generalize across individuals. Effective generalization of these findings to practical task contexts also requires knowledge of appropriate presentation parameters—such as the presentation rate—to maximize processing efficiency without compromising accuracy. To address these issues, we turned to Bayesian search theory.

1.2 Bayesian Search Theory

Statistically optimal search theory was developed during World War II, by the US Navy's Anti Submarine Warfare Operations Research group to optimize the use of limited search resources for a range of search problems: from rescue missions, to locating the wreckage of aircraft and ships. This work, based on maximizing expected utility using a Bayesian framework, has been very influential and continues to form the theoretical basis for modern search operations. The focus of the work reported here has been to apply techniques informed by search theory to estimate the expected performance of a human observer for a given class of targets. We start with a brief review of optimal search theory to illustrate the importance of the detection function, which is the cornerstone of the proposed estimation process. The following explanation is a simplification of the normative theory of optimal search presented by Koopman [3] and Stone [8].

For the sake of simplicity, we assume a discrete search space, i.e., that the targets could be located at one of K locations and that the prior probability of a target at the i^{th} location is given by p_i. The effectiveness of a search process is often determined in part by the search resources that can be allocated to different search locations. Examples of resources that can be distributed include time or the resolution at which the search area is traversed (e.g. coarse vs. fine search grid) under the constraint by which the total resources are limited. The goal of the optimization is to determine the best distribution of search resources over locations. This computation requires knowledge of the relationship between specific resources allocated to locations and the probability of detecting targets, if present. To express this mathematically we denote the resources allocated to the i^{th} location by h_i and define a detection function u(h) as the conditional probability of finding a target at that location, given that the target is at the i^{th} location by the detection function: . Since the resources available to search in most situations are limited, the goal of the search optimization process is to allocate resources to maximize expected detection performance, for example, the number of detected targets,

$$E\{Utility\} = U\{\vec{r}, \vec{h}\} = \sum_{i=1}^{K} u_i\,(h_i)\,p_i, \tag{1}$$

where $\overrightarrow{h} = \{h_1, h_2, ...h_n\}$ is the vector of allocated resources constrained by

$$H = \sum_{i=1}^{K} h_i \tag{2}$$

and $\overrightarrow{r} = \{p_1, p_2, ...p_n\}$ is the vector of prior probabilities Assuming that the detection function u, has certain convexity properties, it is possible to determine the optimal resource allocations by constrained optimization of Equation (1). In particular, using the method of Lagrange multipliers, we can show that the optimal allocation of resources to the i^{th} location must obey the following

$$\frac{d}{dh_i} u(h_i) = \frac{\lambda}{p_i} \tag{3}$$

where λ is a Lagrange multiplier, typically calculated using the constraint expressed by Equation (1). The result in Equation (3) means that the resource allocated to a location should be proportional to the instantaneous rate of detection, combined with prior probability.

In the case of the RSVP search technique outlined here, the resource to be optimized is the time allocated to processing each chip. The idea is to make the search as efficient as possible without compromising detection accuracy. The detection function is, therefore, a fundamental component that will allow us to estimate the tradeoff between efficiency and accuracy. Detection functions are typically empirically derived because of the complex interaction between a target and features of the search area [9]. The next section illustrates a parametric approach to estimation of detection functions.

2 Method

As mentioned above, the estimation of an empirical detection function is of central importance in search theory. In many search domains, proxy targets are embedded in specific search contexts. Empirical search performance data is used to construct functions that help approximate the probability of detecting targets under various scenarios. For example, proxy hikers wearing different types of clothing may be placed in terrain where search and rescue missions are common [9]. Detection performance is assessed as a function of various combinations of search resources and parameters (for example, search using airplanes or helicopters from different altitudes). Fits of psychophysical functions to this data can help answer questions about the optimal combination of resources and parameters to search for the target of interest.

We adapted this approach to characterize detection performance as a function of the RSVP presentation rate based on empirical data. The targets used in our study consisted of surface-to-air missile (SAM) sites (Figure 2). We chose SAM sites because they represent a common type of target that a large population of image analysts (military) are trained to detect. Additionally, they have a well-defined set of visual features that are easy to describe to both experienced image analysts and inexperienced participants. SAM sites also vary considerably in complexity—some targets have features that are prominent enough to pop out with exposures of a few tens of milliseconds, while detection of other SAM sites requires careful reasoning based on prior knowledge and contextual information.

We employed an RSVP experimental paradigm in order to develop an empirically derived detection function. Images were presented to users in blocks that lasted two-seconds. Half the blocks contained targets drawn from the 105 targets extracted from several broad area satellite images. To assure that each image was given equal exposure for processing, a pattern mask followed each stimulus chip. The presentation duration of each image varied between 25, 50, 100, 150, and 200 milliseconds. Ten participants recruited from the population of researchers at Honeywell Laboratories were asked to respond to each target with a button press.

Fig. 2. Example of surface to air missile sites

3 Results

The behavioral data gathered in conjunction with the RSVP study described above were summarized in terms of the probability of detection at each image presentation rate. These estimates correspond closely to the definition of the detection function in the search theory framework. The detection function in this case was similar to a psychometric function in a standard signal detection paradigm. In contrast to a standard psychophysical function, the objective signal strength was not known and its effect had to be estimated from the data. The analyses were based on the assumption that the detection function can be well-approximated by parametric fits to a Gumbel distribution of the form:

$$Pr\{Correct\ Detection \mid T\} = u_i(T) = \exp(-e^{B_0+B_1T}), \tag{4}$$

where B_0 and B_1 are constants to be estimated and T is the presentation rate. In our study, these constants, estimated on data averaged over subjects, were directly related to the objective strength of the target. This family of functions was selected in order to capture long tails of the detection function for some targets. Instead of reporting these abstract quantities, however, it seemed more natural to report the presentation rate at which the detection function reached a fixed point, e.g., $u_i(T_{75}) = 0.75$. These thresholds are then used to represent the target difficulty.

Figure 3 illustrates a few examples of psychometric functions empirically derived from the SAM site data. Each point (blue circle) on the four plots summarizes detection performance averaged across subjects at a particular rate. The green curve fitted through these points is a psychometric function that describes detection performance as a function of performance rate. The vertical red line in each plot represents the detection threshold, i.e. the rate at which detection performance reaches 0.75. The images and associated plots in Figure 3 show how the detection threshold rises as discriminating features of SAM sites (radial arrangement of missiles and launchers, service paths, central RADAR etc.) become less prominent. As described below, detection thresholds estimated in this way provide a way to make inferences about numerous issues that could influence detection performance—including the optimal presentation rate for a given class of targets and individual differences in detection performance.

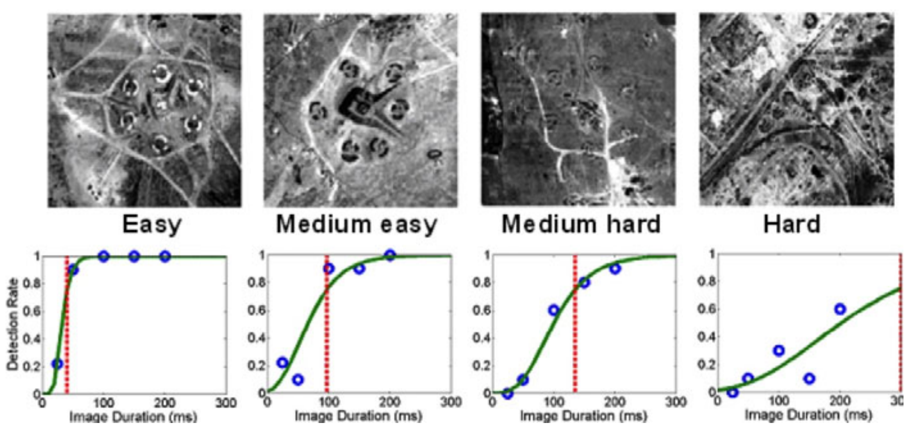

Fig. 3. SAM sites of varying complexity with associated detection thresholds. Detection threshold (red line) estimated from performance data rises as a function of target complexity.

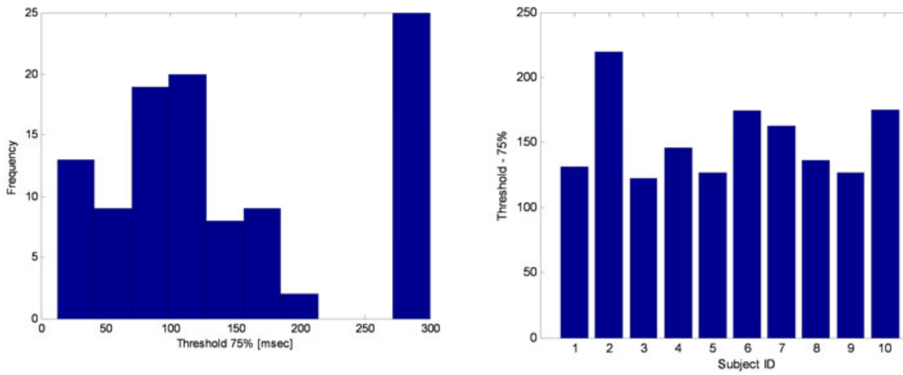

Fig. 4. Distribution of detection threshold across all targets included in study (left). Distribution of average detection threshold across individuals (right).

To assess the overall efficacy of the RSVP technique for the targets of interest, we pooled detection thresholds associated with each of the 105 targets included in the experiment. A histogram based on these thresholds (Figure 4, left) suggests that over 75% of targets could be detected with an accuracy of 75% or more at rates of 200 ms/image or less. Approximately 25% of targets could not be detected with 75% accuracy at the presentation rates used in our study. It is likely that a substantial proportion of the targets that could not be detected are perceptible at rates above 200 ms; however, a subset of these targets may not be well-suited for RSVP-based triage. Whether or not detecting 75% of targets is acceptable depends on the nature of the missions—in particular, based on considerations that balance the risk of missed targets against processing efficiency. Nevertheless, these results suggest that the RSVP modality can provide the basis for efficiently and accurately detecting complex targets of practical relevance.

Our analysis also examined individual differences in performance by pooling average detection thresholds for each participant across presentation rates. Results show that the detection threshold for six of the ten participants was150 ms or lower (Figure 4, right). Nine of ten participants had average detection thresholds below 200 ms. Only one participant had a detection threshold exceeding 200 ms. This result has several implications. First, detection performance is consistent across individuals—suggesting that the RSVP modality may be a viable image screening modality for a wide range of users. Second, empirically derived detection functions can provide a principled way to optimize presentation rates for an individual or a group of individuals. Third, detection functions can also provide a way to screen individuals who may be more effective than others at detecting targets using the RSVP modality.

4 Discussion

The research described here builds on previous work demonstrating the potential to enhance the efficiency of image analysis by combining the RSVP presentation modality with response modalities ranging from EEG-based event related potentials to pupil and motor responses. The analysis described here employs Bayesian search theory to investigate the extent to which these results generalize across a complex class of targets. Additionally, we demonstrate how an empirically derived detection function, a central component of optimal search theory, can provide a principled basis for estimating the optimal presentation rate for an individual or a group and to identify individual differences in search efficacy. The analysis presented above suggests that most SAM sites can be detected at rates of 200ms/chip. Additionally, we find a fairly consistent distribution of detection thresholds across subjects.

The work described here has focused on a detection function based on a single parameter: RSVP presentation rate. However, this approach could be extended to include other parameters that affect search performance, from the spatial scale at which an image is processed to the eccentricity of targets within each image chip.

The described detection function is just one component necessary to optimize search. Optimal search theory emphasizes the importance of considering prior probabilities of targets; search resources, characterized by empirically derived detection functions, should be allocated to regions in proportion to the prior target probability associated with specific regions of the search space. Additionally, the presentation rate for a given area of an image should be determined not just by the detection function, but a joint consideration of the prior probability of a target in the region and the cost associated with the risk of missing a target, if present.

Acknowledgements. This work was supported by the Defense Advanced Research Projects Agency under contract N10PC20048. The views, opinions, and/or findings contained in this article/presentation are those of the author/presenter and should not be interpreted as representing the official views or policies, either expressed or implied, of DARPA or the Department of Defense. The authors thank James Carciofini and Karen Feigh for their contributions to this work. Satellite images included in this paper were produced by DigitalGlobe Inc., Longmont, CO 80501, USA. (c) 2003

References

1. Chase, W.G., Simon, H.A.: Perception in Chess. Cognitive Psychology 4, 55–81 (1973)
2. Huang, Y., Erdogmus, D., Pavel, M., Mathan, S., Hild, K.E.: A Framework for Visual Image Search Using Single-Trial Brain Responses. Accepted by the Journal of Neurocomputing
3. Koopman, B.O.: Search and Its Optimization. The American Mathematical Monthly 86(7), 527–540 (1979)
4. Mathan, S., Whitlow, S., Erdogmus, D., Pavel, M., Ververs, P., Dorneich, M.: Neurophysiologically Driven Image triage: a Pilot Study. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2006, pp. 1085–1090. ACM, New York (2006)
5. Parra, L., Christoforou, C., Gerson, A., Dyrholm, M., Luo, A., Wagner, M., Philiastides, M., Sajda, P.: Spatiotemporal Linear Decoding of Brain State. IEEE Signal Processing Magazine 25(1), 107–115 (2008)

6. Qian, M., Aguilar, M., Zachery, K.N., Privitera, C., Klein, S., Carney, T., Nolte, L.W.: Decision-Level Fusion of EEG and Pupil Features for Single-Trial Visual Detection Analysis. IEEE Transactions on Bio-medical Engineering 56(7), 1929–1937 (2009)
7. Sajda, P., Gerson, A.D., Philiastides, M.G., Parra, L.C.: Single-Trial Analysis of EEG During Rapid Visual Discrimination: Enabling Cortically-Coupled Computer Vision. MIT Press, Cambridge (2007)
8. Stone, L.D.: Theory of Optimal Search. Academic Press, New York (1975)
9. Koester, R.J., Cooper, D.C., Frost, J.R., Robe, R.Q.: Sweep Width Estimation for Ground Search and Rescue. US Department of Homeland Security, United States Coast Guard Operations (G-OPR), Washington DC (2004)