

# UAV See and Avoid Systems: Modeling Human Visual Detection and Identification

Andrew B. Watson  
NASA Ames Research Center, Moffett Field, California

The FAA seeks to characterize the ability of UAV viewing systems to support target detection and identification. Existing system evaluation methods require expensive and time consuming subjective experiments. We hope to replace those experiments with the Spatial Standard Observer, a simple model of human detection and discrimination. This report describes progress on two elements of this project: simulation of an existing subjective data set using the Spatial Standard Observer (SSO), and development of a web-based application for demonstrating SSO-based visibility calculations. Preliminary results indicate the utility of both elements.

## Introduction

The FAA seeks to compile and review the characteristics and performance of existing optical/digital viewing systems that could be used to enhance the human UAV operator's ability to see-and-avoid potential conflicts with other manned and unmanned aircraft. The systems will be characterized by their performance characteristics: field-of-view, field-of-regard, modulation transfer function, focal point, and lens quality, as well as bandwidth and compression. This comparison will be used to determine the ability of these systems to allow detection of static images of differing sizes, at a range of distances in, variety of visibility conditions, i.e., sense-and-avoid.

In this context there is a need to supplement the Army's target acquisition model with a human vision model to predict observers' probability of detection and recognition of aircraft and other targets. In the current Army target acquisition model, these tasks are associated with particular values of N50 for particular image sets and classes, which are obtained by expensive and time consuming subjective experiment. We propose to create and evaluate a tool for computing N50 from a given image set and given classifications, thus obviating the need for subjective measurements. The predicted N50s would be entered in the Army's target acquisition performance model, Night Vision Thermal Imaging System Performance Model (NVTherm), to determine the effects of camera field-of-view, camera field-of-regard, camera modulation transfer function, opposing aircraft size, contrast, distance,

and atmospheric conditions on observers' detection and recognition of an aircraft[1].

We have developed a model called the Spatial Standard Observer (SSO) that allows predictions of visual detection and discrimination of foveal spatial targets (Watson & Ahumada, 2004). The goal of this project was to assess the feasibility of using the SSO to compute N50 values for target image sets.

The first effort in this project has been to simulate the results of a recent psychophysical experiment that estimated N50 for a set of military vehicles[2]. A second concurrent effort has been the development of a prototype tool for calculation of the visibility of manned or unmanned aircraft under specified viewing conditions.

## Target Identification Model

Here we describe the development and evaluation of a model to predict image and object identification. We begin with a description of the experiment whose data will be modeled.

### *Psychophysical Experiment*

The experiment has been more extensively described in another report[2]. Here we provide a brief summary. The experiment consisted of two parts, using visible and infrared imagery respectively.

In each part of the experiment, the source images consisted of 144 digital images, of 12 "objects" in 12 "aspects." An illustration of two of the objects and three of the aspects are shown for the visible and infrared imagery in Figure 1. Each

object is a particular military vehicle, and each aspect is a view of that vehicle. The twelve aspects are approximately the same from vehicle to vehicle. Of the twelve aspects, eight are views from an elevation of seven degrees, while the remaining four are from 0 degrees.

These source images were blurred with Gaussian kernels of 6 possible scales,

$$G(\mathbf{x}) = \text{Exp}\left(\frac{-\pi|\mathbf{x}|^2}{\text{scale}^2}\right) \quad (1)$$

The scales ranged from 5 to 30 pixels in steps of 5. This yields a total of  $6 \times 144 = 864$  images for each image set (visible or infrared). The six

levels of blur are illustrated in Figure 2.

Identification experiments using trained human observers were run separately on each level of blur. Each observer viewed a subset of 144 images of one type (visible or infrared), consisting of 2 aspects for all 12 objects in all 6 blurs. The two aspects were chosen in a quasi-random fashion. The observers were previously trained on identification of these vehicles, using different images. On each trial, the observer attempted to identify the object. The percent correct was recorded. The results are shown in Figure 3.



Figure 1. Example images. Two objects (rows) and four aspects (columns) are shown for both the visible and infrared image sets. The last aspect shows an example of the 0 degree elevation.



Figure 2. Examples of the six levels of blur applied to one image of each type (visible and infrared).

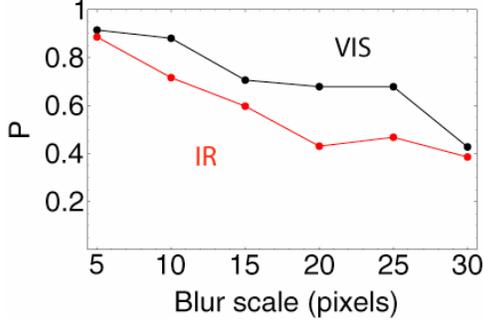


Figure 3. Percent correct identification as a function of blur scale for visible and infrared targets.

### Model

The first model we have considered is a simple image classification machine operating on the basis of a normalized correlation matching rule. This model computes a set of  $N$  discriminant functions, where  $N$  is the number of possible images (in this case,  $N = 144$ ). One discriminant corresponds to each candidate image, and the model selects the image with the largest discriminant.

The matching is assumed to occur in a “neural image” space, which is reached by transforming the image. The transformation consists of a conversion to contrast and filtering by a contrast sensitivity filter (CSF). The CSF is derived from our Spatial Standard Observer (SSO), a simple model of foveal contrast detection[3].

The templates consist of the transformed images. If the presented transformed image is written  $s$  (for sample), then the discriminant for image  $i$  is given by

$$d_i(s) = s \mathbf{g}_i \quad (2)$$

where  $t_i$  is the normalized template. It is not necessary to divide by the norm of  $s$ , since it is the same for all discriminants.

Each transformed image can be expressed as a product of its normalized form and its energy

$$g_k = e_k t_k \quad (3)$$

Thus if image  $k$  is presented,

$$s = e_k t_k + n \quad (4)$$

where  $n$  is a neural noise image (noise in the neural image space). Then

$$\begin{aligned} d_i(s) &= (e_k t_k + n) \mathbf{g}_i \\ &= e_k t_k \mathbf{g}_i + n \mathbf{g}_i \end{aligned} \quad (5)$$

We can divide through by  $e_k$  without changing the ranking of the discriminants,

$$\begin{aligned} d_i(s) &= t_k \mathbf{g}_i + \frac{n \mathbf{g}_i}{e_k} \\ &= \rho_{i,k} + \frac{n \mathbf{g}_i}{e_k} \end{aligned} \quad (6)$$

where  $\rho_{i,k}$  is the correlation (dot product) between each pair of neural images.

If the noise is white and normally distributed with standard deviation  $\sigma$ , then the second term in this expression will be a normally distributed random variable with standard deviation  $\sigma/e_k$ . So finally, each discriminant will be a normal random variable distributed as

$$d_i(s) = \text{Normal}\left(\rho_{i,k}, \frac{\sigma}{e_k}\right) \quad (7)$$

To simulate performance of this model, we simply pick a noise  $\sigma$ , and generate  $N$  discriminant values for a number of trials  $T$  for each of  $N$  sample images. On each trial, the image selected is the largest discriminant, and from these results we can compute percent correct (we can also generate confusion matrices). We compute both percent correct image identification and correct object identification. The performance of the model is controlled by a single parameter:  $\sigma$ , the standard deviation of the “neural noise” added to the sample neural image. In Figure 4, we plot the percent correct for image identification and object identification for images blurred by 30 pixels.

As expected, increasing noise reduces performance. The red and green lines in the figure show the asymptotic guessing performance expected given the numbers of images and objects, and the larger values of noise reach these asymptotes.

Another question of interest is whether the image and object identification performance can

be related by a simple guessing model: is the object identification performance what would be expected by assuming that if the model does not pick the correct image, that it then guesses among the other images. In that case the percent correct object identification ( $P_o$ ) can be computed from the percent correct image identification ( $P_I$ ) as

$$P_o = P_I + \left(1 - P_I\right) \frac{N - 1}{N^2 - 1}. \quad (8)$$

This prediction is shown by the gray curve in Figure 4. Clearly, in this example, the object identification is better than would be expected from this prediction. We call this the "object advantage" (OA). The OA is negligible at 5 pixels blur, but increases to a max of about 0.13 at 30 pixels. Without an aperture (see below), it is about the same for VIS and IR. With an aperture, it is smaller for IR than for VIS. Possible sources for the OA are: background (without aperture), object color (for visible), and overall object size. We will return to this point later.

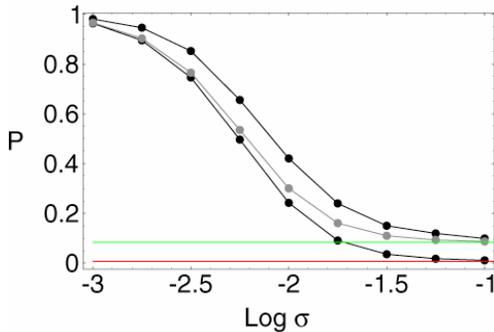


Figure 4. Percent correct image (lower black curve) and object (upper black curve) identification for various levels of the noise standard deviation. These results are for visible targets at blur scale = 30 pixels. Green and red lines indicate predicted guessing performance. The gray curve is object identification predicted from image identification using a guessing model (see text).

### Object Identification vs Blur Scale

The results for image identification can also be plotted as a function of blur scale, as shown in Figure 5. The value plotted is percent correct object identification (as in the upper curve in Figure 4), and each curve is for a different noise sigma.

The figure also includes (blue and red curves) the data from the human observers. No attempt has been made at this point to find the best fitting value of noise  $\sigma$ , but it is clear that a value of around -2.25 yields a rough approximation to the human data for visible images, and -2 for infrared images.

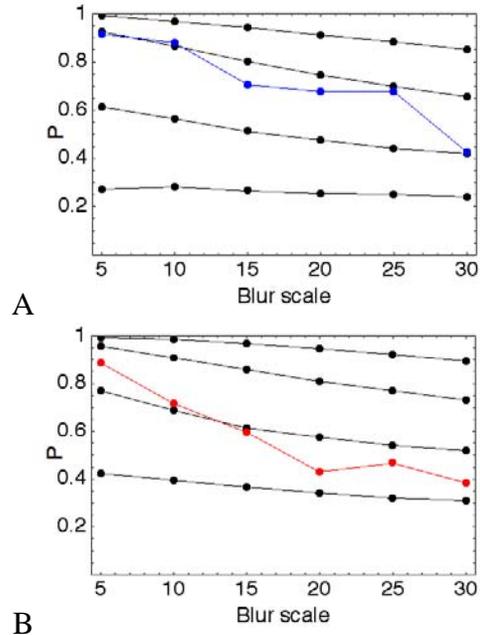


Figure 5. Simulated percent correct object identification as a function of blur scale for several different values of neural noise ( $\text{Log } \sigma = -2.5, -2.25, -2., -1.75$ ). The blue and red curves are the human data. A) visible, B) infrared.

### Removing the Background

As noted above, object identification performance is better than expected from the guessing model, which indicates that on average different aspects of one object are more similar (as images) than are aspects of another image. This could be due in part to the object background, which is nearly constant from aspect to aspect. To test this we have computed results for images with the background removed. Aperture images defining the object area were provided by the U.S. Army Night Vision and Electronic Sensors Directorate. The apertured image was constructed as  $\text{image} * \text{aperture} + 2048 * (1 - \text{aperture})$ . An example of the construction of one apertured image is shown in Figure 6.

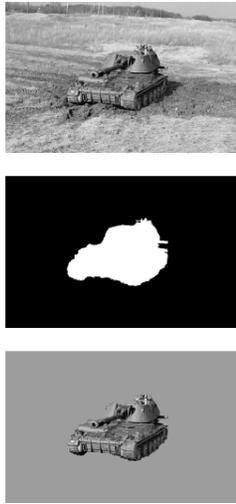


Figure 6. Construction of an apertured image. A) Original image, B) aperture, C) apertured image.

The model results obtained using the apertured images are shown in Figure 7. Overall, performance is somewhat better than for the original images. The visible image performance for  $-\text{Log } \sigma = -2.25$  is now closer to the data, while the infrared data lie between  $\text{Log } \sigma = -2.5$  and  $-2.25$ .

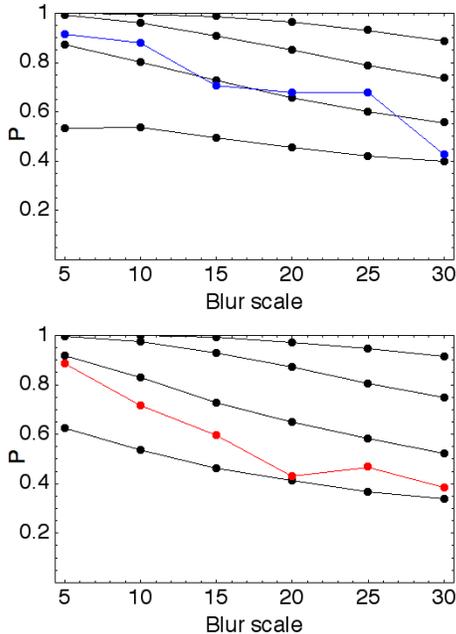


Figure 7. Object identification performance vs blur scale for apertured images. Details as in Figure 5.

### Visible vs Infrared

One purpose of the original psychophysical experiment was to determine the relation between N50 for visible and infrared images of similar objects. If the N50s were the same, that would allow the same metric to be used regardless of the image type. However, in that experiment the estimated N50s differed by about 50% (7.5 visible, 11.5 infrared)[2].

Figure 8 compares model results for visible and infrared. A short summary is that performance is somewhat better for infrared than for visible, but that this advantage largely vanishes with apertured images. Recall that human performance is slightly lower for infrared, so this constitutes a small discrepancy between model and data.

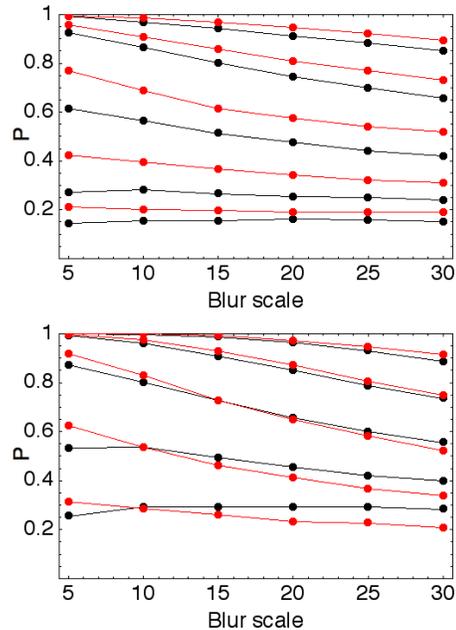


Figure 8. Object identification performance vs blur scale for visible (black) and infrared (red) images. A) Original, B) apertured. Other details as in Figure 5.

### Summary

A very simple identification model incorporating the Spatial Standard Observer can generate performance similar to human data for both visible and infrared imagery. Some discrepancies remain, notably the slightly steeper decline with blur, and the poorer performance with infrared

imagery, found in the human results. We hope to investigate these matters further in the second stage of this project.

Future work on this part of the project will include alternative SSO-based models, as well as other human data sets[4]. We hope to understand better the reasons for infrared vs visible performance. We also want to work with aircraft rather than tank images.

## Visibility Calculator

In a second part of this project, we have begun development of a prototype application to predict visibility of aircraft targets as they might be seen from a UAV. Conversely, the tool could be used to predict visibility of the UAV from another aircraft. A screen shot of the prototype application is shown below.

The tool allows the user to select an aircraft, as well as various viewing parameters. The tool then computes the visibility of the aircraft, expressed in units of JND. The tool is currently online and operational at the URL shown in the figure.

The tool operates by computing a rendered image from a selected 3D model. The rendered image is then processed using the current version of the Spatial Standard Observer (SSO). The tool is implemented using webMathematica, an extension of the Mathematica language[5]. The current version of the prototype is only a proof of concept, and must be augmented by realistic optical and atmospheric effects, and must be calibrated in both geometric and photometric aspects. We plan to accomplish these augmentations in the second phase of this project.

## Acknowledgments

I thank Albert J. Ahumada, Jr. and Jeffrey B. Mulligan for helpful discussions, and thank Ron Driggers and Eddie Jacobs of the Army Research Lab for providing the target images and background information on the psychophysical experiment.

## References

1. Vollmerhausen, R.H. and E. Jacobs, *The Targeting Task Performance (TTP) Metric A New Model for Predicting Target Acquisition Performance*, Modeling and Simulation Division Night Vision and Electronic Sensors Directorate U.S. Army CERDEC Fort Belvoir, VA 22060.
2. Driggers, R.G., et al., *Fifty-percent probability of identification ( $N_{50}$ ) comparison for targets in the visible and infrared spectra*, US Army Night Vision and Electronic Sensors Division (NVESD).
3. Watson, A.B. and A.J. Ahumada, Jr., *A standard model for foveal detection of spatial contrast*. Journal of Vision, in press.
4. Moyer, S., et al., *Cycle Criterion for Fifty-Percent Probability of Identification for Small Handheld Objects*. 2005, US Army CE-COM, RDEC Night Vision and Electronic Sensors Directorate Ft. Belvoir, VA.
5. Wolfram, S., *The Mathematica Book*. 5th ed. 2003, Champaign, IL: Wolfram Media.

See and Avoid

http://eveleth.arc.nasa.gov:8080/webMathematica/abw/saa.jsp

# NASA "See and Avoid" Visibility Calculator

Aircraft:

Background (0-1):

Ambient lighting:

Light source distance:  Color:

Distance:

Viewpoint:



View of the selected aircraft under the specified viewing conditions. Click and drag on the image to rotate, shift-click-drag vertically to zoom.



View with addition of cloud background.



View from specified distance.

Visibility = 16.5641 JND.

Uncalibrated Prototype.

Date = {2005, 9, 16, 17, 5, 32.354173}

Mathematica Version = 5.2 for Mac OS X (64 bit) (June 20, 2005)

Figure 9. Screen shot of web-based visibility tool.