The Targeting Task Performance (TTP) Metric

A New Model for Predicting Target Acquisition Performance

Richard H. Vollmerhausen
Eddie Jacobs

Modeling and Simulation Division
Night Vision and Electronic Sensors Directorate
U.S. Army CERDEC
Fort Belvoir, VA 22060

Technical Report AMSEL-NV-TR-230

Approved for public release, unlimited

This page intentionally left blank

Contents

Preface / iv

1. Introduction / 1
2. History of Target Acquisition Modeling / 9
3. Modeling Human Vision / 18
    - 3.1 Contrast Threshold Function / 20
    - 3.2 Contrast Threshold Function in Noise / 22
    - 3.3 Visual Bandpass Filters / 24
    - 3.4 Model Calibration / 27
    - 3.5 Validity of Weber's Law / 28
4. Contrast Threshold Function of an Imager / 30
    - 4.1 Effect of Blur on Imager CTF / 31
    - 4.2 Effect of Contrast Enhancement on Imager CTF / 33
    - 4.3 Effect of Display Glare on Imager CTF / 34
    - 4.4 Limit on Modulation Gain / 35
5. Definition of Target Acquisition Tasks / 38
6. Predicting Target Acquisition Performance from Imager CTF / 43
    - 6.1 Predicting Probability versus Range / 44
    - 6.2 Meaning of Model Probabilities / 46
    - 6.3 Field Test Example / 47
    - 6.4 Estimating Task Difficulty (V50) / 48
7. Modeling Sampled Imagers / 52
    - 7.1 Response Function of a Sampled Imager / 53
    - 7.2 Impact of Sampling on Range Performance / 54
        - 7.2.1 Discussion / 56
8. Modeling Reflected-Light Imagers /57
    - 8.1 Staring Focal Plane Arrays /58
        - 8.1.1 Interlace /62
        - 8.1.2 Snapshot and Frame Integration / 63
    - 8.2 Direct View Image Intensifiers / 64
    - 8.3 I^2 Optically Coupled to CCD or CMOS / 66
    - 8.4 CCD or CMOS Array inside I^2 Tube / 68
    - 8.5 Predicting Probability versus Range / 69
        - 8.5.1 Contrast Transmission through the Atmosphere / 69
        - 8.5.2 Effect of Contrast Enhancement / 71
        - 8.5.3 Calculating Probability of Task Performance / 73

8.6 Minimum Resolvable Contrast / 73
9. Modeling Thermal Imagers / 75
9.1 Signal and Noise in Thermal Imagers / 77
9.2 CTF_{sys} for Thermal Imagers / 79
9.3 Predicting Probability versus Range / 81
9.3.1 Contrast Transmission through the Atmosphere / 81
9.3.2 Effect of Contrast Enhancement /82
9.3.3 Calculating Probability of Task Performance /83
9.4 Minimum Resolvable Temperature / 84
References / 85
Appendices / 91
A. Metric Validation Experiments / 91
B. Experiments with Low Contrast and Boost / 99
C. Recognition Experiment / 102
D. Experiment with Laser Speckle / 105
E. Predicting Component MTF and Other Details / 114

Preface

This report describes a new target acquisition performance model which uses the Targeting Task Performance (TTP) metric. Like its predecessor, the famous Johnson criteria, the new model assumes that range performance is proportional to image quality. Simplicity of implementation is therefore maintained. However, the TTP model predicts image quality in a different fashion. In addition to overall better accuracy, the TTP metric can be used to model sampled imagers, high frequency boost, non-white noise, and other features of modern imagers which cannot be accurately modeled with the Johnson criteria.

The Johnson criteria are used almost universally to predict range performance. Johnson uses the resolving power of an imager as a metric of sensor "goodness" for target acquisition purposes. For a given target to scene contrast, resolving power is the highest spatial frequency passed by the sensor and display and visible to the observer. He multiplies the resolving power of the imager (in cycles per milliradian) by the target size (in milliradians) to get "cycles on target." Johnson published a table of the number of cycles on target needed to detect, recognize, identify, or perform other target acquisition tasks; these are his "criteria" for target acquisition. The basic assumption underlying the Johnson metric is that all electro-optical imagers are the same in some broad sense. The performance of the imager can be determined solely by the highest spatial frequency (f_J) visible at the average target to background contrast. When the Johnson method works, it is not because f_J is important per se, but rather because an increase in f_J represents an improvement in the contrast rendition at all spatial frequencies. However, with sampled imagers, f_J is more an indicator of sample rate than image quality. Also, because the Johnson metric is based on the system response at a single frequency, it cannot predict the effect of tailoring the image frequency spectrum through digital processing. For example, the benefits of edge sharpening by high frequency boost cannot be predicted.

In Appendix A of this report, the predictions of the Johnson criteria are shown to be fundamentally flawed due to its insensitivity to imager characteristics below the limiting frequency. This flaw makes predictions for many modern imaging systems inaccurate. Experimental data show the problems with the Johnson criteria and illustrate the robust performance of the TTP metric. The simplicity of implementing a range performance model with the Johnson criteria is retained by the new metric while extending applicability of the model to sampled imagers and digital image enhancement.

The new target acquisition model includes another fundamental change, also described in this report. The current models to predict Minimum Resolvable Temperature and Minimum Resolvable Contrast were introduced in 1995. Products like the NVTherm thermal model and the SSCAM (Solid State Camera) TV model differ from their predecessors because the contrast limitations of vision are incorporated into these models. Incorporating the eye contrast limitations allows the modeling of image intensifiers, TV, and sensitive thermal imagers which were previously not accurately modeled. However, the 1995 model set continued the use of the "matched eye filter." Since this filter does not

reflect psychophysical reality, those models are only accurate when the noise is spectrally flat (white) compared to the signal. Digital image processing, particularly high frequency boost or image restoration, can lead to a distinctly non-white noise spectrum. The modeling of modern imagery requires a change in the eye filters.

Because the Johnson-based models have been widely used for so long, considerable attention is paid in this report to the history and assumptions underlying the older model. In addition, the new TTP model is described in detail. Also, the theory predicting human contrast threshold when using an EO imager is thoroughly documented.

1

Introduction

In Figure 1.1, the soldier is using an imaging sensor, hoping to quickly identify whether the tank is a threat. This report describes a model which predicts the probability that he correctly identifies the target. The problem is tackled in two parts. First, the soldier's quality of vision when using the sensor and display is quantified. Most of the report is devoted to this topic. Second, the relationship between quality of vision and performing a visual task, such as identifying the tank, is discussed.

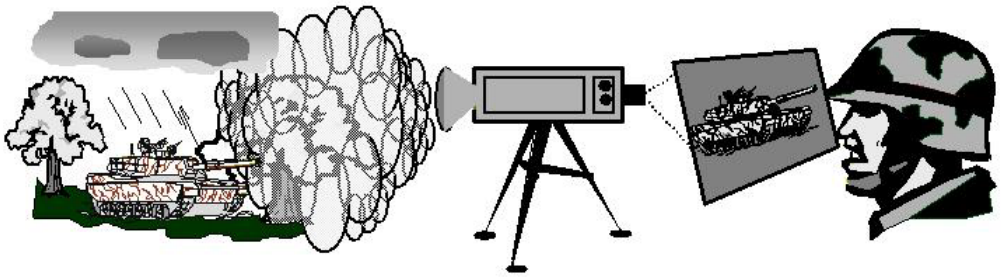


Figure 1.1 Soldier Trying to Identify a Tank as Friend or Foe

The theory in this report is couched in terms of the observer viewing the world through the imager. The imager extends the observer's vision because it provides advantages over human eyesight. The target can be magnified; that is, the angle subtended by the target at the eye can be greatly increased, making it easier to see. The imager also lets the observer see light outside the visible wavelengths, often a great advantage because the target signatures are more robust. There is more night illumination at near infrared wavelengths than the visible, for example, so that image intensifiers work better in that spectral band. Another example is thermal imagery, which does not depend at all on natural illumination. On the negative side, however, the imager blurs the target and adds noise to the viewed scene.

The degradation due to imager noise and blur are in addition to the natural limitations of human eyesight. If the imager were perfect—no blur from the optics, detector, or display and no noise from the photo-detection process—the observer's range performance would still be limited by his vision. Image quality results from the inherent limitations of human vision in combination with imager blur and noise. The limitations of human vision depend, in turn, upon the display luminance and contrast.

The most widely used measures of image quality are v*isual acuity* and *resolving power. Visual acuity* has the connotation that high contrast (black on white) letters or symbols are used to check vision. The observer who reads the smallest letters has the best visual acuity. With sensors, the term *resolving power* has the same connotation. Bar patterns are generally used for imagers; the best imager displays the smallest bar pattern. Although

commonly used and easy to test, these high-contrast measures do not adequately quantify how well a person can see with the naked eye or through the imager.

A scene consists of many luminance levels. The eye achieves an integrated view of objects by connecting lines and surfaces. These lines and surfaces do not share a particular brightness throughout their extent. For example, the background immediately behind a target might not be uniform, and yet the eye sees a full or partial silhouette. Perspective is gained from converging lines which might vary in both sharpness and luminance with increasing range. Slight changes in hue or texture can provide an excellent cue as to the distance and orientation of an object and possibly indicate the nature of the surface characteristics. Acute vision requires the ability to discriminate small differences in gray shade, not just the ability to discriminate small details which happen to have good contrast.

In Figure 1.2, the picture of Goldhill has an average modulation contrast of 0.22. The 3-bar charts to the right have contrasts of 0.04, 0.08, 0.16, and 0.32 with average luminance equal to the average of the picture. When noise is added and the picture blurred, as shown at the bottom of Figure 1.2, high contrast details are still visible, but low contrast details disappear. This is illustrated by the bar charts at the bottom which were degraded in the same way as the Goldhill picture. A quantification of visual performance requires that resolution be measured for all shades of gray in the image. The means of achieving this quantification is described later.

Figure 1.2. Picture of Goldhill and 3-bar Charts of Various Contrasts Measuring resolution for the average or peak contrast does not adequately quantify picture quality.

The model proposed here is more complex than resolving power; hopefully the need for this added complexity will become apparent as the report proceeds. For the present, we

quote Lucien Biberman who is quoting G.C. Brock (Chapter 8 in Biberman, 1973; Brock, 1965).

> "Before we can make progress in the use of our new techniques it will be necessary to bypass two obstacles, the first of which is the existence and firm establishment of resolving power, and the second is the belief that science will give us one number quality index that will supplant all previous evaluation techniques.
>
> "Resolving power has been in use for so long that it has come to be thought of as something fundamental which determines other aspects of image quality and has some very special significance. Whenever a new criterion of image quality is proposed, we at once ask 'How does it relate to resolving power?' instead of considering it in more general terms. And because resolving power is used for so many different purposes, and gives a one number answer, it is assumed that any new technique must be inferior if it does not do the same. As we have already seen, resolving power can serve many purposes because it does not serve any of them well."

The imager model must account for both hardware characteristics and human vision. In EO imagers, blur, noise, and contrast all limit our ability to see details. Further, unless the display is big and bright, the physiological limitations of the eye cannot be ignored. A picture might appear grainy when presented at high display luminance and not noisy at all when presented at low display luminance. This does not mean that the picture is better in some quantitative sense when presented at low display luminance; our inability to see the noise infers an equivalent inability to see contrast gradations within the scene itself. Hardware characteristics do not, by themselves, establish image quality. Rather, hardware characteristics interact with human vision to establish how well the scene is perceived through the imager.

Depending on scene conditions and sensor control settings, the dominant hardware factor limiting performance can be blur, noise, or contrast. Blur results primarily from factors like diffraction or aberrations in the objective lens and summing of the incoming light over the detector instantaneous field of view. Summing the light from different points in the scene results in the blurring of scene detail. Noise is generally associated with the photo-detection process. In the theoretical limit, signal is proportional to the number of photo-electrons generated in the detector. Noise is proportional to the square root of the number of photo-electrons. Contrast can be degraded by the atmosphere. For example, sunlight scattered by the atmosphere into the sensor line-of-sight can seriously degrade contrast. Contrast can also be degraded by the glare of ambient light reflecting off the display or by improper display settings. Blur, noise, and contrast limit our ability to see detail and therefore limit our ability to identify targets or to discriminate between target and background.

Figure 1.3a is a thermal image of a tank. The tank has been exercised, and the road wheels and engine are hot, giving the tank a thermal signature which is distinct from the background. In Figures 1.3b and 1.3c, the tank is viewed from progressively greater distance. Optical magnification is used so that the tank appears to be the same size, but

diffraction in the objective lens has blurred the tank's details. Noise is not visible in the image; the tank is difficult to identify at the longest range because of the blur. In Figure 1.3d, the tank has cooled off. In order to see the tank, the gain on the imager is increased. Increasing imager gain in Figure 1.3e makes the tank visible, but also makes the detector noise visible. In Figure 1.3f, noise associated with photo-detection obscures the tank image.

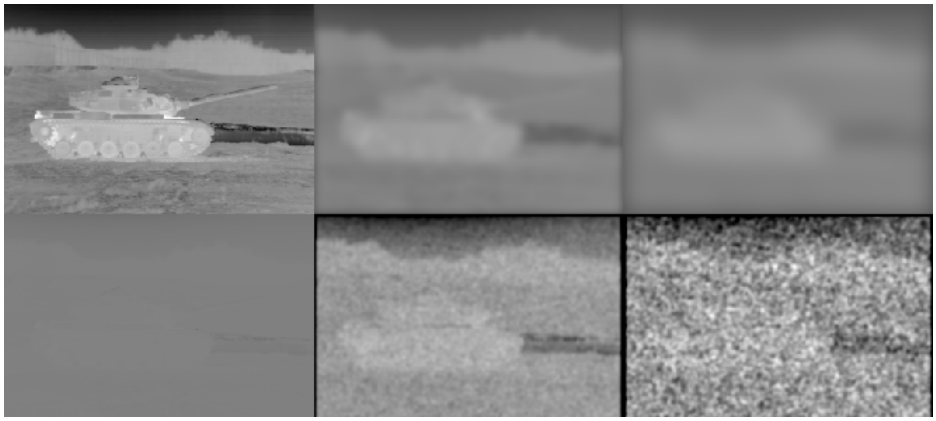


Figure 1.3 Thermal Image of a Tank Showing Effects of Blur and Noise
At top, pristine image in (a) is blurred by imager in (b) and further blurred in (c). At bottom, pristine but low-contrast image in (d) becomes noisy images in (e) and (f) because increased gain amplifies detector noise.

Blur and noise also affect the performance of reflected light sensors. Generally, performance is limited by blur or contrast under good illumination conditions and by noise under poor illumination. This is because, in the theoretical limit, signal to noise is proportional to the square root of photo-current. As illumination decreases, photo-current decreases, and noise becomes more dominant.

Figure 1.4a shows a visible band image of a tank. In Figures 1.4b and 1.4c, the tank is viewed from progressively greater distance. Optical magnification is used so that the tank appears to be the same size, but diffraction in the objective lens has blurred the tank's details. Noise is not visible in the image; the tank is difficult to identify at the longest range because of the blur. In Figure 1.4d, illumination has decreased and the tank is not visible. In image 1.4e, the camera gain is increased and the tank is again visible, but the low illumination makes the picture noisy. In Figure 1.4f, illumination has decreased to the point that the tank is not visible in the noise.

A third factor important in determining performance of night vision sensors is display contrast, especially when display luminance is less than photopic. In Figure 1.5, the picture of Lena becomes clearer from left to right because contrast increases; neither signal to noise nor blur changes. Contrast limitations are especially important when display luminance is low; the eye's ability to discriminate gray levels in an image degrades as display luminance decreases.

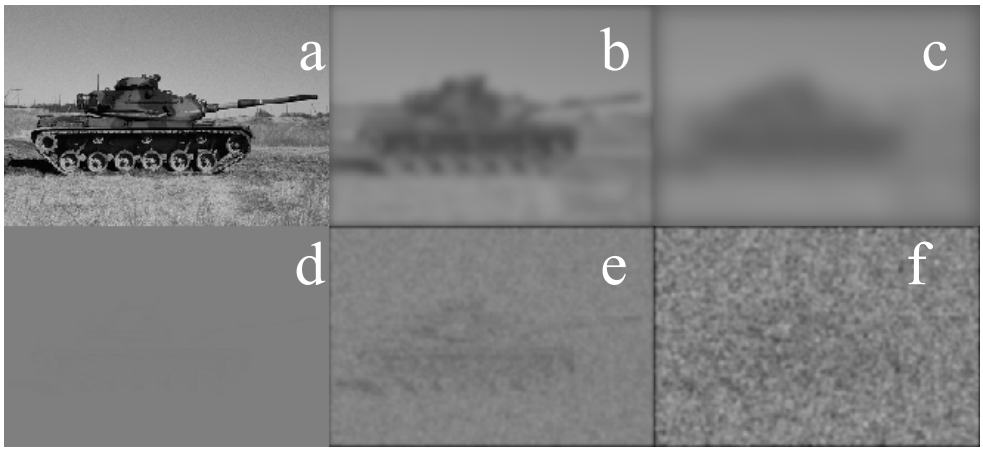


Figure 1.4 Visible Image of a Tank Showing Effects of Blur and Noise

Figure 1.5 Picture of Lena Showing Contrast Increasing from Left to Right

Low display luminance might occur because of imager limitations. For example, due to insufficient light gain, early image intensifiers provided less than 0.01 foot Lamberts (fL) eyepiece luminance under starlight scene illumination. (10 fL is considered low photopic luminance.) Early attempts to model image intensifier performance failed because the model was based on signal to noise in the image. However, at the low display luminance, neither the signal nor the noise was clearly visible to the observer. Image intensifiers were not accurately modeled until the contrast limitations of the eye were incorporated into the model.

Low display luminance is not uncommon. Display luminance might be low because the operator chooses to maintain dark adaptation. During night flight, for example, military pilots flying without goggles set instrumentation displays to between 0.1 and 0.3 fL; this permits reasonable viewing of the instruments while maintaining dark adaptation in order to see outside the aircraft. Regardless of the reason for a non-optimized display, the result is degraded human performance when using an imager. It is common and even typical for the display luminance of a night vision device to be less than a foot Lambert, and this low display luminance is an important factor in determining the performance of the night vision imager.

All four factors affecting performance—the blur, noise, and contrast of the imager as well as the physiological limitations of the eye in adjusting to a non-optimized display—must

be handled by the model. All four factors affect the targeting performance expected from the imager. For both reflective and thermal imagery, performance is generally limited simultaneously by a combination of factors. That is, the image is not less blurred just because noise is present.

The theory in this report covers all types of EO imagers. Imagers of reflected light like sunlight or starlight operate in the spectral band between 0.4 and 2 microns. Thermal imagers sense emitted light (that is, heat). Thermal imagers operate in the mid-wave infrared (3 to 5 microns) or the long-wave infrared (8 to 12 microns). These spectral bands are defined by atmospheric "windows" with good transmission. The units used to describe signal and noise for thermal imagers are different than the units used when modeling reflected light sensors. However, aside from the details of calculating signal and noise, the basic target acquisition theory is exactly the same. In both cases, the observer is looking at a display of the blurred and noisy image of a target. The model predicts the effect of blur, noise, and display characteristics on target acquisition task performance.

That is not to say that interpreting thermal imagery is as easy as understanding a picture in the visible spectral band. For most people, imagery becomes progressively harder to interpret as the wavelength increase from visible to near infrared to short-wave infrared. Thermal imagery, which is emissive rather than reflective, is very difficult to interpret for the untrained observer. However, the difficulty of the observer's task is included in the target acquisition model, not in the image quality model. The same image model is used for all imagers.

Traditionally, thermal scenes are characterized with absolute, blackbody temperature differences, and thermal imager frequency response is measured with 4-bar patterns. Illuminated scenes are characterized by contrast, and the frequency response of reflected light imagers is characterized with 3-bar patterns. An absolute temperature difference in the scene can, of course, be converted to a contrast, just as a contrast can be algebraically converted to an absolute illumination difference. The main difference in the historical treatment of thermal and reflected light imagers is that the two are characterized using different bar patterns.

In this report, all imagers are treated the same. The choice between absolute differences in the scene or scene contrast is arbitrary. However, as discussed in Part III, contrast is normally used to characterize the eye. The use of contrast when modeling EO imagers simplifies the presentation of the theory. Further, it is customary to use sinewave patterns for eyeball measurements, and the use of sinewaves is consistent with our sensor model. Fourier theory is used to model system blur and noise. The development of a target acquisition metric is made easier by characterizing human vision with sinewaves; this allows easy integration of the eye behavior into the Fourier frequency domain model.

It is understood that sinewave measurements are not practical in the laboratory. However, there is a known relationship between bar chart response (either 3- or 4-bar) and sinewave response. These conversions are described where appropriate in the theory sections on specific imaging technologies.

Most of the report is devoted to predicting how well the observer can see through the imager. Our ultimate goal, however, is to predict how well the observer can detect, recognize, or identify targets. To meet that goal, an image quality metric is needed as a link between quality of vision and task performance.

The Johnson criteria are used almost universally to predict range performance based on sensor resolution. Johnson proposed that an imager's utility for target acquisition purposes was proportional to its resolving power (Johnson, 1958). That is, for a given target to scene contrast, the highest spatial frequency passed by the sensor and display and visible to the observer determines the probability that an observer can correctly identify a tactical vehicle or perform other visual discrimination tasks. He used his limiting-resolution metric to establish criteria for target acquisition tasks.

Johnson performed some engineering experiments using image intensifiers to simultaneously view bar-charts and scale models of tactical vehicles. He published a table giving the required "cycles on target" for a 0.5 probability of detecting, recognizing, identifying, and other levels of target discrimination. "Cycles on target" is the imager's bar resolution in cycles per milliradian multiplied by the angular subtense of the target in milliradians. Johnson used the target's critical dimension to determine its angular size; critical dimension corresponds, more or less, to the minimum of the target's height or width as viewed by the sensor. D'Agostino later substituted the square root of viewed area for critical dimension, and updated the cycle criteria needed for target discriminations (Howe, 1993).

The Johnson metric uses limiting bar-chart resolution as an indicator of sensor goodness for target acquisition purposes. Predictive accuracy of this metric is best when comparing "like" sensors and conditions. The metric is not compatible with many features found in modern sensors. For example, it is not compatible with sampled imagers. Further, the Johnson metric fails to predict the impact of frequency boost on range performance.

The basic assumption underlying the Johnson methodology is that all electro-optical imagers are the same in some broad sense. The performance of the imager can be determined solely by the limiting resolution frequency (f_J) visible at the average target to background contrast. When the Johnson criteria work, it is not because f_J is important per se, but rather because an increase in f_J represents an improvement in the contrast rendition at all spatial frequencies. However, with sampled imagers, f_J is more an indicator of sample rate than image quality. Further, as pointed out persistently by Fred Rosell, the Johnson metric fails to accurately predict the effect of noise on task performance. The observer appears to require more sensor resolution when the resolution is noise limited as opposed to spatial frequency response limited (Rosell, 1979 & 2000).

The desired approach to modeling sampled imagers is to incorporate a targeting metric that does not have the problems associated with the Johnson metric. Work on a replacement metric started several years ago (Vollmerhausen, 07/2000 and Driggers, 2000). This report describes how the new TTP (Targeting Task Performance) metric is calculated and used. The logic of this metric is discussed by Barten; TTP is similar to Barten's SQRI (Square Root Integral), except that linear rather than logarithmic integration is used (Barten, 1999). It is also similar to van Meeteren's Integrated Contrast Sensitivity (Task, 1976 and Tannas, 1985). A variety of experiments were performed

showing the problem with the Johnson criteria and illustrating the robust behavior of the new TTP metric (Vollmerhausen, 2003 and Appendix A).

The organization of this report is outlined as follows. The next two sections provide needed background. Part 2 is on model history; this section discusses the assumptions upon which models of the last half-century were based. Part 3 discusses some of the remarkable properties of human vision and describes how vision is characterized in our model. Part 4 describes how the hardware characteristics of the sensor and display are combined with the limitations of the eye to form a model of threshold vision through an imager. The target acquisition tasks predicted by the model are defined in Part 5. Part 6 describes how an image quality metric is used to relate the quality of threshold vision through an imager to the probability of acquiring a target at range. Part 7 discusses how sampled-image artifacts affect performance. Parts 8 and 9 present details on the models for reflected-light and thermal imagers, respectively.

Appendix A summarizes ID experiments which show the problems with the Johnson criteria and the robust performance of the TTP metric. Experiments included both thermal and visible imagery. Further, experimentation was done with well sampled images and a variety of MTF and noise, poorly sampled imagers, and imagers with high frequency boost and colored (spectrally non-uniform) noise. Appendix B discusses experiments run with very low contrast targets. Appendix C describes a recognition experiment. This experiment was done for two reasons. First, to check the TTP metric with a task difficulty easier than target identification. Second, the experiment checked that the sampling range adjustment is correct for target recognition. Appendix D describes an ID experiment where the images were corrupted by laser speckle. The experiment is significant because laser speckle has a very non-uniform power spectrum; the imagery is highly corrupted with low frequency, very high contrast noise. Appendix E provides some details needed to implement the target acquisition model.

2

History of Target Acquisition Modeling

Electro-optics (EO) technology has flourished since World War II; even a brief mention of all the important contributors and events would require volumes. The present discussion is focused on human-in-the-loop target acquisition models, and only the major threads of model development history are followed.

The history of modeling EO imagers traces back almost 60 years to the pioneering work of Otto Schade. In the introduction of his four-part paper "Electro-Optical Characteristics of Television Systems," Schade noted that the standard which must be met by an EO image was established by the capabilities and optical characteristics of the eye (Schade, 1948). He pointed out, for example, that the visibility of "grain" fluctuations decreases with brightness, so a comparison of the signal to noise characteristics of different imaging technologies should be made at the same display brightness level. The following is a quote from the conclusion of Part IV.

> "The quality of television and photographic images depends in a large measure on three basic characteristics of the imaging process: the ratio of signals to random fluctuations, the transfer characteristic, and the detail contrast response. These characteristics are measured and determined by objective methods which apply equally well to all components of photographic and electro-optical imaging systems." He states that hardware can be rated on an objective numerical basis, and then continues: "An interpretation of the numerical values obtained by calculation or measurement of the three characteristics that determine image quality requires correlation with the corresponding subjective impressions: graininess, tone scale, and sharpness. This correlation has been established by analyzing the characteristics of vision and by including these characteristics in an evaluation of the over-all process of seeing through an image-reproducing system."

In 1956, Schade published a model of the eye (Schade, 1956). The visual system was treated as an analog camera; performance of the camera was quantified using sinewave response, contrast sensitivity, and other psychophysical data. Schade combined the physical data on hardware and psychophysical data on human vision and created a holistic model of the observer's aided vision. Schade postulated that, for each retinal illumination, information transfer could be calculated by a knowledge of threshold signal to noise ratio and signal transfer characteristics. Over-all transfer characteristics were obtained by integration of intensity steps and by considering the sampling efficacy of the rods and cones; this integration of "statistical units" constituted his *passband* metric. He

used this model to compute the degradation in visual performance when the imager was inserted between the scene and eye.

> "One of the objects for constructing an analog [of the eye] is its use for obtaining visual evaluations for image characteristics by calculation, to eliminate subjective observations. This calculation is done by computing the degradation in visual response when an external process is inserted between the object and the eye. The degradation in resolution, for example, is given by the ratio of two line numbers obtained at a given small response factor; one with the eye alone, and the other for the eye in cascade with the external imaging process. The total degradation may be rated by the logarithm of the ratio of the equivalent passbands [the normal visual passband and the combination eye-imager passband]."

Schade's work provided fundamental and widely accepted design guidelines for television and other EO systems. However, Schade's sensor performance model was complex and difficult to adapt to changing conditions. Although his model was widely studied, it was not widely used. To our knowledge, the ability of Shade's analog eye model to predict target acquisition performance was never assessed. However, based on the form of the passband model, our experiments indicate that it would not be a good predictor of target acquisition performance. Shade later simplified his passband metric to include only integration over the sensor MTF (Shade, 1973). The simplified version of the passband metric was evaluated by Task and did not accurately predict target acquisition performance (Task, 1976; Tannas, 1985).

Meanwhile, the model that was eventually used by virtually everyone for the next fifty years was being developed by Coltman (1954). Coltman developed a model to predict the resolving power of fluoroscopes. Richards adopted Coltman's model to predict the resolving power of night vision imagers (Richards, 1967). Johnson postulated that target acquisition performance using an imaging sensor was proportional to the resolving power of the imager (1958). A modified version of the Coltman/Richards model for the imager and the Johnson model for predicting target acquisition range were brought together by Ratches, Lawson, and others in the Night Vision Laboratory Static Performance Model (Ratches 1975, 1976, and 2001). The NVL model used Fourier transform theory and communications theory concepts which were fully developed for imaging sensors by Lawson (1971).

Derivatives of Coltman's model are so widespread that the model is generally presented without attribution. The simple assumptions which are the basis for the model are seldom questioned. This is unfortunate. Coltman's focus was fluoroscopy, and his model requires that the display be optimized. He reasoned as follows.

> "The advent of electronic devices for brightening images has made it possible in principle to remove the optical and physiological deficiencies of the eye. In the limit there will remain only the quantum noise inherent in the signal itself."

Coltman based his model on ideas put forward by Barnes and Czerny (1933), de Vries (1943), and fully developed by Rose (1948). Rose assumed that the absorption of luminous flux by photoreceptors of the retina would be accompanied by the same

statistical fluctuations (shot noise) as occurs in any square-law detector. He considered only low light level circumstances where quantal fluctuations could be expected to dominate the detection process. Further, he considered circular disks of sufficient size to be resolved by the eye. Under these circumstances, Rose assumed that the eye would integrate the signal and noise over the disk area. The result predicts Piper's law; for a given adapting luminance, the product of signal threshold and the angular size of the disk is a constant. When compared to experimental data, Rose's theory worked for intermediate sized disks but failed for both small and large disks. The detection of small, circular disks is predicted by Ricco's law; for small objects, the product of signal threshold and disk area is a constant. For large objects, detection occurs at a constant contrast.

Coltman postulated that shot noise in the eye would not be significant compared to the photo-detection noise associated with the fluoroscope. He assumed a big, bright display and that noise in the sensor photo-detection process would dominate the perceptual signal to noise because of the gain provided by the display. Realizing that bar-pattern detection might be mediated by different perceptual processes than circular disk detection, Coltman assumed that the visual system acted as a spatial integrator over an area related to the object to be detected and admitting noise from the same area. He did not assume signal and noise integration over a single bar. Finally, he followed Rose's assumption that, for a detection to occur, a constant signal-to-noise ratio threshold must be achieved at some point in the visual processing chain. In Figure 2.1, the eye is summing both signal and noise over an area related to the bar size. Once the integrated signal exceeds the noise by a fixed threshold, the observer can differentiate between the bar and the space.

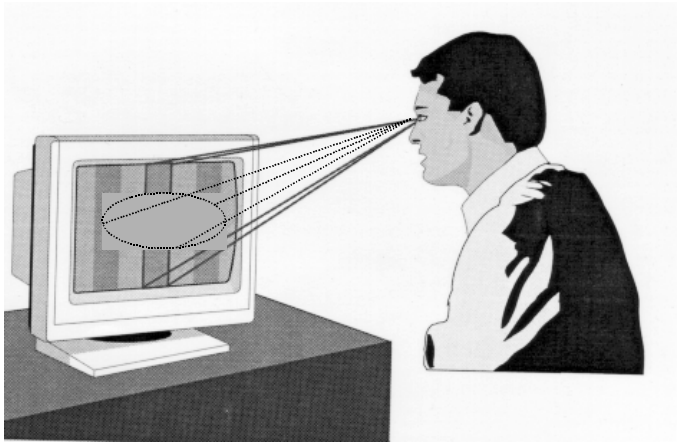


Figure 2.1 The Eye as a Spatial Filter
Coltman sensor model assumed that the eye integrates signal and noise over some area of the image related to the bar size.

Coltman tested his theory experimentally. Observer variability was too great to conclusively demonstrate the validity of his assumptions, but neither did the data indicate that his assumptions were in error. Most analysts accepted the tenets put forward in the Coltman model. That is, the signal contrast needed to detect a bar pattern varied in proportion to the square root of bar area. In the presence of noise, large bars were easier to see than small bars.

Coltman did not postulate that the eye was integrating over a single bar of the pattern; he could not determine the actual shape or size of the area being integrated. In Coltman's experiment, the eye could be using any fraction or multiple of the bar or bars to establish signal to noise. The signal to noise ratio threshold (SNRT) required by the eye to see the bar is an experimental result. Increasing integrated area by a factor of four reduces SNRT by a factor of two. Since SNRT is not known independent of the experiment, the integration area cannot be predicted. By the same logic, for white noise, the *shape* or *spatial weighting* of the integrated area cannot be predicted. The nature of the spatial filter was not established by Coltman's experiment.

Richards (1967) adopted Coltman's theory to model night vision devices. He simplified Coltman's equation, and made it appear more definitive, by assuming that the eye integrated over the area of a single bar. Coltman explicitly included an arbitrary multiplier that flagged the ambiguous nature of his results. In simplifying Coltman's equation, Richards set the arbitrary multiplier equal to one. This theory, that the eye is integrating over the bar area, later became known as the "matched filter" model.

Experiments like those of Coltman and later Rosell (Rosell, 1973 and Rosell, 1979) demonstrate that the eye filters noise, but do not definitively establish a filter function. The calibration constant (SNRT) adapts the model to any shape and placement of the filters in the frequency domain, providing that bandwidth is proportional to bar frequency and that the noise is white.

The matched filter model was simple and seemed to explain observed behavior. The eye's remarkable ability to see objects in noise has been experienced by many engineers over the years; this lends credence to the idea that the eye is spatially integrating over the object being viewed. This model became the basis of the NVL (later the Night Vision and Electronic-Sensors Directorate or NVESD) performance models until 1995. In the 1975 to 1995 model, the eye acted as a matched filter, integrating signal over the bar area, and admitting noise from the same area. The bar was detected (threshold reached) when the peak signal to RMS noise ratio exceeded a fixed value (SNRT) independent of bar size. The noise arose solely from the detector; as detector approached zero, so did predicted threshold.

The NVL model did include a pupil-dependent eye MTF factor that was added to account for vision limitations; the model also included a factor representing temporal signal integration by the eye that depended on display luminance. These factors were added to overcome the assumption of an "optimized" display. However, pupil dilation plays a minor role in luminance adaptation. These additions did not change the fundamental nature of the model; the contrast threshold limitations of the eye were ignored.

The models predicted Minimum Resolvable Temperature (MRT) for thermal imagers and Minimum Resolvable Contrast (MRC) for imagers of reflected light. However, these models were only accurate for some imagers. Early thermal imagers, for example, were noisy and had sufficient gain that the noise itself could generate a photopic or near photopic display luminance. These sensors met the assumptions laid down by Coltman; the display could be sufficiently optimized that the imager was detector noise limited.

Other technologies, however, could not be modeled. Early attempts to model image intensifiers failed, because the eyepiece luminance of the device was low mesopic. The display could not be optimized, and eye limitations could not be ignored. Further, the performance of day sensors could not be modeled. Daylight illumination provided plenty of signal and detector noise became insignificant; performance was contrast limited. Since the early NVL models were strictly based on a signal to detector noise calculation, contrast limited situations were not correctly modeled.

Alternative assumptions about the nature of the eye filter received some attention. Sendall and Rosell proposed substituting the synchronous integrator model (Sendall, 1979; Rosell, 2000). However, under practical conditions, the predictions of the matched filter model and the synchronous integrator model differ only slightly (Lawson, 1979). Overington (1976) proposed using the signal and noise associated with the boundary rather than the area. He suggested that gradients in the contour are important and should be weighted by their visibility. The static predictions in British Aerospace's Orcale Model use these concepts, but the details of the Oracle Model implementation have not been published to our knowledge.

It was recognized by a number of researchers that the Coltman model ignored fundamental limitations of the eye. Schnitzler (1973) modeled the "noise-required input contrast" of a displayed target by cascading the quantal limitations of the EO imager and eye. Overington paid a great deal of attention to the workings of the eye, emphasizing the presence of noise and blur both external and within the eye. He proposed an *equation of vision* which was a function of the threshold intensity difference divided by adapting luminance (the psychometric contrast). Object detection depended upon intensity gradients in the displayed image, with the gradient spacing defined by eye receptors and gradient amplitude scaled by the psychometric contrast. Overington provides alternate formulas for small, intermediate, and large objects and for the effect of the blur associated with visual aids. However, he does not model the effect of system related noise.

Because of the failure of the "standard" model to predict image intensifier performance, model development continued at NVL also. The work was probably done by Kornfeld and Lawson using ideas put forward by van Meeteren (1986), but the available working papers are not signed and do not cite references. In this model, "eye noise" is assumed to be proportional to the contrast threshold function of the eye. The eye noise is root-sum-squared with the signal to noise term used in the Ratches model. This addition to the NVL model was significant because the modified model provided correct results in the limit of zero detector noise. That is, as the system noise decreased to zero, the eye became the limiting factor. However, this addition made little difference in model predictions; image intensifier predictions were still quite inaccurate. An image intensifier model was eventually published by NVL; the IIV4 Image Intensifier model was published in 1995. However, that model used empirical fits to laboratory data in an attempt to correct the problems with the theory.

One concept fundamental to all of the above theories is that the signal was detected when it exceeded the shot noise by a fixed amount. The noise was sometimes only the shot noise associated with the sensor photo-detection, and sometimes the shot noise was

modeled as the combined noise from sensor and eye neural noise. The signal might represent detecting a bar or circular disk against a bland background; in this case, the models were called Minimum Detectable Temperature or Minimum Detectable Contrast. When calculating Minimum Resolvable Temperature and Minimum Resolvable Contrast, the signal was the bar-space-bar modulation of a bar pattern. Whether the model was predicting the presence of an object in noise or detection of bar modulation, both types of model employed the same assumptions. First, the visual system integrated over the bar or simple object. Second, the SNRT was constant regardless of the size of the object or bar pattern. Third, the eye noise, when considered, was associated with primary photo-detection by the eye; therefore, the eye noise was proportional to the square root of display luminance.

The history above has focused on the detection of simple, circular disks or bar-patterns through and imager. An equally important factor in target acquisition is relating the detection of simple patterns to the process of interpreting real, complex images. Although the Johnson criteria is used almost universally, many alternatives have been proposed.

Rosell used the matched filter concept to calculate sensor resolution; however, he felt that the Johnson criteria range predictions were imprecise (Rosell, 1979; Rosell, 2000). The Johnson metric tends to be optimistic when the image is noisy. That is, more "cycles on target" are needed to perform an acquisition task when the imagery is limited by noise rather than blur. Rosell suggested adjusting the Johnson range predictions based on the signal to noise established by target contrast at range and the sensor's noise equivalent temperature difference. The resulting range model was somewhat clumsy to implement. The validity of Rosell's criticism was widely understood, however, and alternatives to the Johnson model were pursued by Rosell, Biberman, and others for many years (Biberman, 2000). The model by Roberts, Biberman, and Deller is described here as an example.

A fixed resolution on the target is selected; this is the cycles across target to achieve a 0.5 probability of task performance. For each range, the known target size and required number of cycles across the target yields a spatial frequency. The MRT curve is used to find the threshold contrast needed to resolve that frequency. A signal to noise ratio is formed based on target apparent contrast and the MRT threshold contrast. The probability of task performance is then based on that signal to noise ratio. However, according to well documented but unpublished evaluations by Lawson and Johnson, these alternatives never proved as successful as the Johnson criteria in estimating field performance. Models based on Rosell's concept tend to predict very high probability out to the range where the Johnson criteria would predict 0.5 probability. At that range, probability drops abruptly to zero. While it has been argued that this is realistic for poor weather, clear weather predictions follow the same trend. A sharp drop in acquisition probability is not observed in practice.

There are a number of MTF-based measures of image quality. The Johnson metric is one of these, as are Modulation Transfer Function Area, Integrated Contrast Sensitivity, Square Root Quality Index, and many others. The Target Task Performance (TTP) Metric described later in this report is also an MTF-based measure of image quality. The idea for these metrics started with Shade's equivalent passband. See Task (1976), Tannas (1985), Snyder (1973, 1988), Beaten (1991), and Biberman (1973, 2000/Chapter 22) for surveys

of this area. These metrics share the concept that image quality can be quantified by some weighted integral of signal modulation which exceeds the eye contrast threshold. For example, the Johnson frequency is defined by the spatial frequency range over which the apparent target contrast exceeds the eye threshold. For the other metrics, the amount that the signal modulation exceeds threshold at each spatial frequency is important. All of these metrics share the virtue that range prediction is easily implemented; in every case, range is simply proportional to the value of the metric.

Researcher's in the field have found that, in general, MTF-based metrics account for more than half the variance in performance across the various displays tested. Although the correlation between a particular metric and performance varies greatly from experiment to experiment and task to task, limiting resolution measures like the Johnson metric are generally among the worst performers. However, prior to the TTP metric, experiments at NVL have shown the Johnson criteria to perform better than Modulation Transfer Function Area, Integrated Contrast Sensitivity, and other metrics evaluated (Vollmerhausen, 07/2000).

The discrepancy in experimental conclusions appears to be based on the form of the analyses. Most researchers change one or more calibration "constants" to fit calculated metric values to experimental data. They argue that changes in task, observation conditions, and observer-to-observer physiology requires that the metric be uniquely adapted to each experiment. From the standpoint of a target acquisition model, however, such a procedure cannot be used to predict performance; the procedure requires experimental data on which to base a fit. While all models have one or more calibration constants, those constants must be determined once and then used for all predictions. Under those constraints, the Johnson criteria have proven to be a reasonable predictor of performance, better than other MTF-based metrics like MTFA, ICS, and SQRI.

Overington (1976) and van Meeteren (1990) both theorized that targets are recognized by a process of detecting critical features. This general concept has been the focus of several researchers (Biederman, 1987; O'Kane, 2000). The van Meeteren model, as summarized by Vos and van Meeteren (1991), will be described as it is the most complete in terms of predicting range performance.

Target acquisition is determined by a process of detecting characteristic details. The size, contrast, and number of characteristic details visible to the observer determines the probability of acquisition. Each detail is treated as a circular disk with detection based on a Minimum Detectable Contrast model. In van Meeteren's model, eye noise is represented as a fixed fraction of the contrast threshold at each luminance level. Detection of the critical detail is based on the contrast signal exceeding the quadrature combined detector and eye noise by a fixed amount. One continuing problem with these models is the pre-determination of critical features. It is difficult to adapt the model to a new target set.

One important aspect of van Meeteren's work is explicit task definition. In his 1990 JOSA paper, he describes target recognition as choosing an object from a known confusion set. That is, targets are recognized by differentiating them from the possible alternatives. This means that the features which uniquely define a target are those which differentiate that target from others in the set. Most researchers ignore the important step

of defining the experimental task. By not recognizing that all discriminations are comparisons, many researchers fall into the trap of analyzing experimental data one-target-at-a-time. The result is that a model which appears to predict one experiment beautifully fails to predict subsequent experiments. This distinction—that target acquisition models predict the ability to choose one target of a set, rather than predicting the absolute probability of recognizing or identifying a particular target—is particularly important when trying to assess the success of feature-based models.

The logic of critical-feature recognition is intellectually appealing, but a practical model which incorporates target-set features to predict range performance has not been offered. It should be noted, however, that accepting the idea that high level visual discriminations are required to recognize targets does not invalidate image quality models. Accepting an image quality model like the Johnson criteria does *not* infer that we are not looking at internal target features. The target is not in the model; no judgment is being made about what is being viewed.

The inclusion of overall target–set dimensions and average contrast in range predictions tends to confuse this point. Those parameters are used to decrease variance in range predictions. For example, for ship identification, the "critical dimension" is found to be the vertical height of the ships to be discriminated, and this is included in the model. Tactical military vehicles have more observable features with a side view than front view; the road wheels and gun, for example, are better viewed from the side. The current use of square root of target area when making predictions for tactical vehicles has been found to adjust model output in the correct way. In general, larger, higher contrast targets are easier to see, and including these factors in the model decreases variance in the predictions. However, the fundamental concept behind an image quality model is: *see better, see further*. Of course target details are used to recognize and identify targets. Better image quality lets the observer make these discriminations at longer range.

The fluctuation theory developed by Rose provides a limiting criterion for detection under low luminance conditions. The basic assumption, quite correct for very low display luminance, is that liminal vision is established by the shot noise associated with the retinal photo-detection mechanism. In virtually all cases where target acquisition modelers have considered the nature of the eye, they have assumed that shot noise established the significant limitations of eyesight. This is not the case. The target acquisition task is dependent on the characteristic behavior of higher-order visual processing within the brain.

For any practical display luminance, the contrast limitations of the human eye are established by the visual cortex, not the retina. In 1995, the NVL model which predicts threshold resolution versus spatial frequency was modified to account for these contrast limitations (Vollmerhausen, 1995 and 2000; Driggers, 1999). The modified model accurately predicts image intensifier performance over a wide range of scene illumination and eyepiece luminance conditions. Further, as will be discussed, this model does an outstanding job of predicting the results of experiments using thermal, visible, and laser imagery. The model is applicable to the whole range of EO imagers.

However, the 1995 NVL model continued to use the matched filter concept. Like all models using this filter assumption, it can only be used with sensors where the noise is

essentially white (flat) over the frequency spectrum of the signal. While this can be a serious limitation with modern sensors, the limitation was not serious for previous generations of EO imagers.

Prior to the widespread use of sampled imagers and digital processing, one could assume sensor noise to be essentially white in comparison to the signal. The scene was filtered by the optics and detector as well as the electronics, display, and eye. The noise was only filtered by the electronics, display, and eye. The white noise assumption was valid for the vast majority of sensors. It is still true today that the noise in most EO imagers is white compared to the signal spectrum. In modern sensors, however, digital enhancement of the image can make the noise distinctly non-white. An upgrade to the model is needed.

An upgrade to the 1995 NVL model to correct the eye filters is presented in this report. In this model, the matched filters are replaced with bandpass filters. The new eye filters are based on psychophysical data collected over the last three decades. The combination of the new eye filters and the TTP metric provides complete flexibility in modeling modern EO imagers.

3

Modeling Human Vision

In this section, some of the marvels, complexities, and limitations of human vision are described. The nature of the immediate task requires us to focus on limitations. However, recognizing the capability and the resulting complexity of eyesight provides a needed insight: the nature of vision cannot be encompassed in a simple model. As shown in the charts below, eye behavior changes significantly with luminance and with angular eccentricity from the fovea. This explains why the theory in this report treats the human visual system as a "black box." The threshold response of the eye to sinewave gratings is used to characterize vision; this is experimental data collected by psychophysicists.

The eye provides some quality of vision over a billion to one range of scene illumination. To accomplish this, the eye has cones for photopic or daytime vision and rods for scotopic or night vision. The distribution of rods and cones within the eyeball is shown in Figure 3.1. The highest density of cones is at the center of the fovea, called the foveal pit. There are no rods in the foveal pit, a region in the center of the retina about 200 microns in diameter. The foveal pit subtends about a quarter inch on a display viewed from 15 inches.

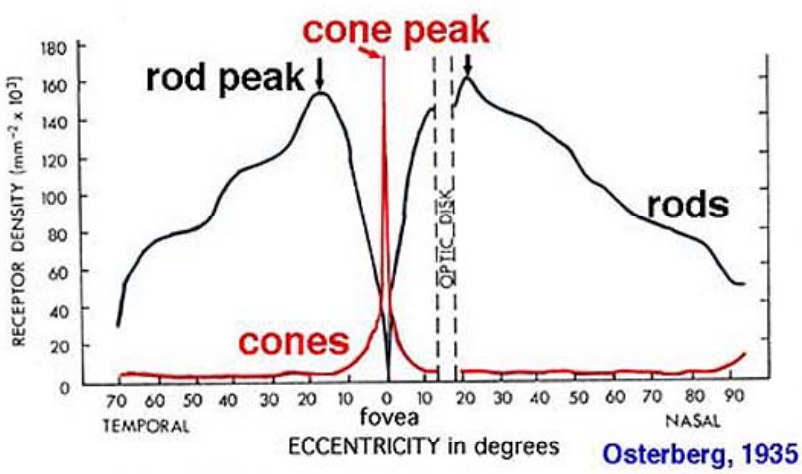


Figure 3.1 Distribution of rods and cones in the retina of the eye.
(Figure courtesy Webvision)

Rods and cones are not equally sensitive to visible wavelengths of light. Unlike the cones, rods are more sensitive to blue light and are not sensitive to wavelengths greater than about 640 nanometers, the red portion of the visible spectrum.

Although factors like retinal processing and pupil dilation play important roles, photo-pigment bleaching is the primary means for adapting both rods and cones to varying

illumination. A threshold versus intensity (tvi) curve can be obtained by testing observers using a small disk of light against a uniform luminance background. When rods or cones are isolated, four sections of the tvi curve are apparent: dark light, Square Root Law (de Vries-Rose Law), Weber's Law, and saturation (Aguilar and Stiles, 1954). Figure 3.2 shows a tvi curve for rod vision. The figure plots the just-visible difference in luminance (ordinate) versus the display luminance (abscissa). "Dark light" is internal, neural noise. The second part of the tvi curve is limited by quantal fluctuation; this is the square root law or de Vries-Rose Law region. The next section of the curve follows Weber's Law; the threshold is a constant fraction of luminance. Given sufficient light, the eye operates on the principle of contrast constancy; this is an important feature of our visual system. In a natural scene, object to background contrast is fairly independent of ambient illumination. The final part of the tvi curve is saturation at high light levels.

According to Ricco's Law, the eye sums quanta over an area. Threshold is reached when the product of luminance and stimulus area exceeds a constant value. In other words, when luminance is halved, a doubling in stimulus area is required to reach threshold. Summation area varies with eccentricity. In the fovea, complete summation occurs over about 0.1 degree. Ricco's Law holds for an area of a half degree at 5° eccentricity increasing to an area of about 2° at an eccentricity of 35° (Davidson, 1990). Spatial summation occurs due to the convergence of photoreceptors onto ganglion cells; clearly, spatial summation limits resolution.

Visual acuity is the greatest at the center of fixation and decreases with eccentricity. See Figure 3.3 for a plot of visual acuity versus eccentricity. There is a close correlation between cone density and visual acuity out to about 2 degrees (Green, 1970).

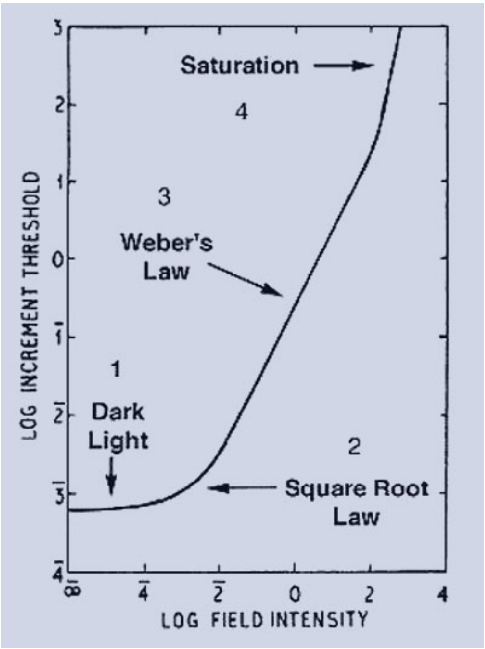


Figure 3.2 Threshold versus intensity curve for rods; similar results are found for cones. (Figure courtesy Webvision)

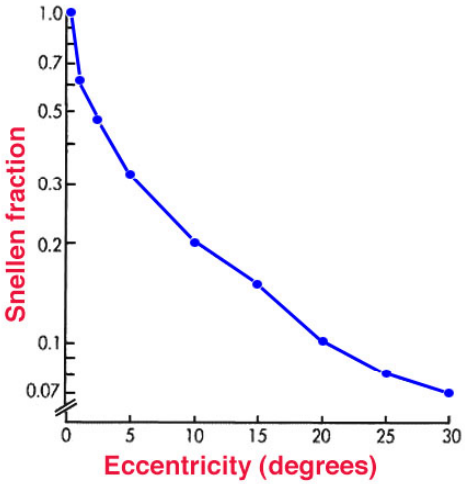
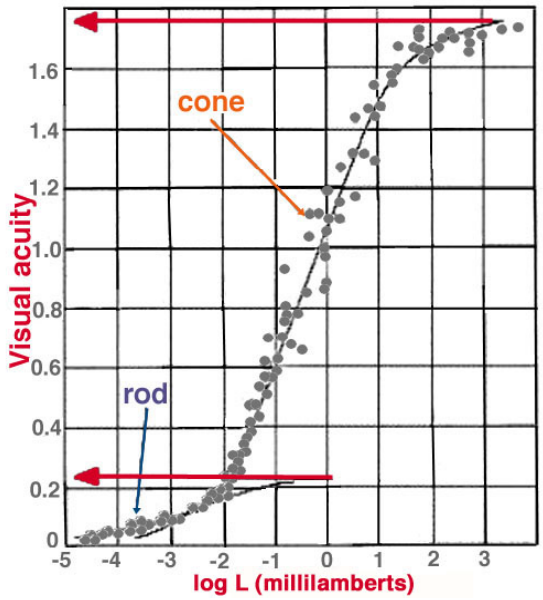


Figure 3.3 Plot of visual acuity versus eccentricity for photopic luminance. (Figure courtesy Webvision)

The illumination range where both rods and cones work together is called mesopic vision. The rods saturate at illumination levels above 10 fL; the cones cease to be important in mediating vision at just below 0.01 fL. The luminance from 0.01 to 10 fL is essentially the range of display luminance used in military night vision systems. Displays used in daylight would be brighter, of course. As shown in Figure3.4, visual acuity varies greatly over the mesopic range of display luminance. (The milliLamberts units used in the figure are almost equal to fL.)

Figure 3.4 Visual Acuity at Mesopic Light Levels (Figure courtesy Webvision)



Although the variation in visual acuity with display luminance has been measured, it is difficult to predict. The interaction between rods and cones is not well understood. Rods and cones are distributed differently over the retina. Rods and cones have different spectral responses, use different photo-pigment chemistry, saturate at different light levels, and employ different neural summation and processing schemes.

The limitations of human vision are important when predicting the targeting performance of an EO imager. However, a reliable theory for predicting visual behavior is not available. In the target acquisition model, experimental data collected by psychophysicists are used to describe human vision.

3.1 Contrast Threshold Function

The Contrast Threshold Function (CTF) is one of the most common and useful ways of characterizing human vision. Objects and their surroundings are of varying contrast. Therefore, the relationship between visual acuity and contrast allows a better understanding of visual perception than acuity measured only with high contrast (black on white) charts.

In Figure 3.5, the observer is viewing a sine-wave pattern. While holding average luminance to the eye constant, the contrast of the bar pattern is lowered until no longer visible to the observer. That is, the dark bars are lightened and the light bars darkened, holding the average constant, until the bar-space-bar pattern disappears. A decrease in

contrast from left to right is shown at top, right in the figure. The goal of the experiment is to measure the amplitude of the sinewave that is just visible to the observer.

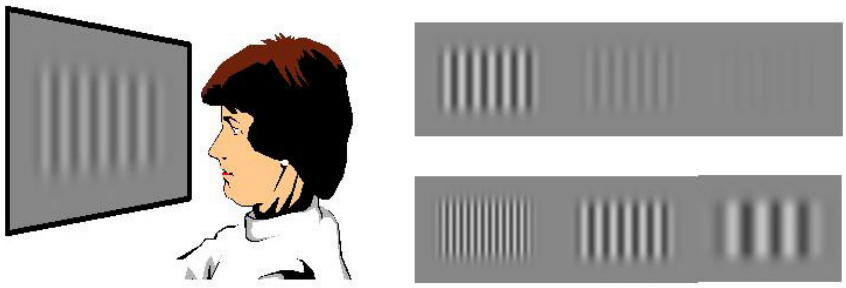


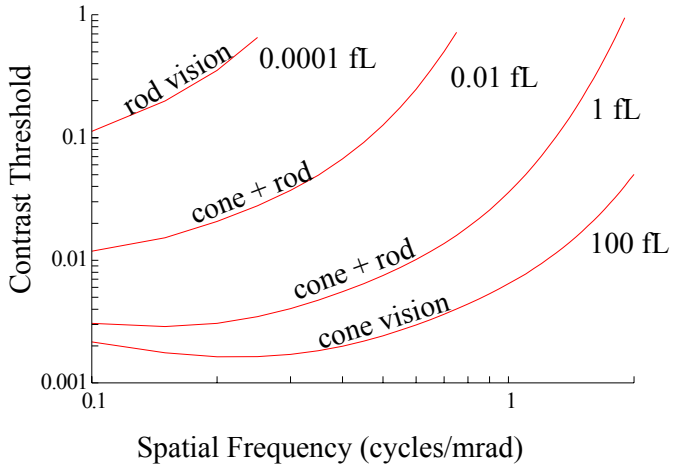
Figure 3.5 Measuring CTF

Most published CTF data is taken with two alternative, forced choice (2afc) experiments. In these experiments, the observer is shown one blank field and one with the sinewave. The observer must choose which field has the sinewave. These experiments measure the sinewave threshold where the observer chooses correctly half the time independent of chance. That is, the 2afc experiment provides the threshold which yields a 0.75 probability of correct choice. The procedure is repeated for various bar spacings—that is, for various spatial frequencies. See the bottom, right of the figure for an illustration of spatial frequency; high spatial frequency is at the left, lower spatial frequency to the right. The curve of contrast threshold versus spatial frequency at each light level is called the CTF at that light level.

Figure 3.6 shows CTF curves for various adapting luminance; the abscissa is spatial frequency and the ordinate is contrast threshold. Each curve shows CTF for a different light level to the eye. Remember that these curves use modulation to describe contrast; that is, contrast equals (bright – dark)/(bright + dark).

Figure 3.6 CTF of Eye at Various Light Levels

Luminance values are in footLamberts.



At each light level, the limiting resolution is the frequency where the CTF curve crosses unity contrast. Limiting resolution provides the smallest detail that will be visible at that light level, and this detail is only visible at the highest possible contrast. CTF provides much more information than limiting resolution; CTF provides the threshold contrast value at all spatial frequencies.

Few real-world objects are totally reflective or totally absorptive; contrast is seldom unity in a real-world scene. A typical scene consists of an infinitude of contrast gradations. The eye's ability to see small contrast differences is critical to quality vision. From the figure, note that the eye loses its ability to see small contrast changes as cone vision is lost. The CTF curve rises as light level decreases. This rise in the CTF curve results in lower limiting resolution and also results in loss of the ability to see small contrast differences at any spatial frequency. An interesting aspect of the CTF curves is that at the higher light levels, people have better threshold vision at middle spatial frequencies than at low spatial frequencies.

3.2 Contrast Threshold Function in Noise

Our model for predicting the effect of display noise on CTF was first described by Vollmerhausen (1995, 2000). This CTF model is currently used in the thermal model and other EO imager models published by the U.S. Army. However, those previous references did not provide a detailed discussion of the CTF model itself. This section starts with some history on modeling CTF in the presence of display noise, describes the van Meeteren model which is used quite often, and then provides a discussion of the current CTF model.

Nagaraja did experiments where he found that noise had an effect on detection threshold which could logically be explained by assuming that the brain was taking the root-sum-square (RSS) of display noise and some internal eye noise (Nagaraja, 1964). In the following equations, CTF_n is measured threshold modulation in the presence of noise, N is display noise modulation (RMS noise divided by twice the display luminance), and κ and N_{eye} are parameter fits.

$$CTF_n^2 = \kappa^2 \left(N^2 + N_{eye}^2 \right) \tag{3.1}$$

Nagaraja then observed that, if N_{eye} and κ are constant, then plotting the square of threshold in noise versus the square of external noise amplitude, a straight line with slope κ^2 and intercept CTF^2 should result. CTF is the measured threshold modulation without external noise.

$$CTF_n^2 = \kappa^2 N^2 + CTF^2 . \tag{3.2}$$

When Nagaraja plotted the experimental data, he found that the plot of N^2 versus CTF_n^2 was linear at 1 fL but that the plots for 0.1 fL and 0.01 fL were not. So Equation 3.1 was correct for 1 fL but was less accurate at lower display luminance. Other investigators have found that Equation 3.1 is approximately true for a wide range of conditions and tasks (Pelli 1981, Legge 1987, Van Meeteren 1988, Pelli 1999). Different tasks include detecting small disks against a uniform background, reading letters, detecting bar patterns, and sinewave threshold detection.

Both κ and N_{eye} are often stated in the literature to be constants fit to the experimental data implying that they are constants relative to both the observer and to the stimulus used in the experiment. If either is changed, both of these factors can change as well. This

means that if an experiment is conducted with several different sinusoidal gratings, κ and N_{eye} will be different for each grating and each observer.

For a limited range of spatial frequency gratings and for photopic luminance, van Meeteren has demonstrated that κ in Equation 3.2 varies slowly; he treats κ as a constant. See Barten (1999) for additional discussion on Van Meeteren's treatment. However, this is the same assumption as used by Lawson to develop the IIV4 Image Intensifier Model; this model does not provide good predictions when display luminance is mesopic, which is almost always the case (Vollmerhausen, 1995).

The current model is derived as follows. From Equation 3.1, at each specific frequency and light level,

$$CTF^2 = \kappa^2 N_{eye}^2 \tag{3.3}$$

$$\kappa^2 = \frac{CTF^2}{N_{eye}^2} . \tag{3.4}$$

Using (3.4) in (3.2)

$$CTF_n^2 = CTF^2 \left(1 + \frac{N^2}{N_{eye}^2} \right) \tag{3.5}$$

or

$$CTF_n^2 = CTF^2 \left(1 + \frac{\sigma^2}{n_{eye}^2} \right) \tag{3.6}$$

where σ is the RMS noise on the display and n_{eye} is the RMS eye noise expressed at the display.

Using Weber's Law, assume that eye noise is proportional to display luminance (L). This proportionality holds over most of the functional luminance range of the human eye (Pelli, 1999; Blackwell, 1958; Section 1.632 of Boff, 1988; Webvision, 2003). For display luminance above the de-Vries-Rose Law region and for statically presented stimuli, the visibility of foveally presented signals is limited by noise arising in the cortex after spatiotemporal and binocular integration (Raghavan, 1989).

$$CTF_n^2 = CTF^2 \left(1 + \frac{\alpha^2 \sigma^2}{L^2} \right) \tag{3.7}$$

It should be remembered that eye noise is a concept used to explain non-zero thresholds. The actual reason that the liminal signal is greater than zero is not known. Rose and de Vries correctly assumed that the statistics associated with photo-detection limits psychometric contrast at low luminance. In this case, signal is proportional to luminance

and noise is proportional to the square root of luminance, so psychometric contrast decreases in inverse proportion to the square root of luminance. However, this assumption only holds for the lowest absolute luminance needed for rod or cone operation. At higher luminance levels, signal detection threshold is proportional to luminance, and psychometric contrast is constant. The reason for this change in behavior as luminance increases might not actually be noise, but rather an adaptation of the visual system to aid the brain in interpreting imagery. Whatever the cause, Figure 3.2 does indicate that threshold is proportional to display luminance over most of the luminance range usable by the eye.

Once the calibration constant (α) is determined by experiment, Equation 3.7 provides an accurate means of predicting the effect of display noise on contrast threshold. As described in the next section, however, some of the frequency spectrum of the display noise is filtered out by the eye. The value of α is given after the noise filter is discussed.

3.3 Visual Bandpass Filters

It is important to note that the signal and noise in Equation 3.7 are taken with respect to the bandpass properties of the human visual system. In other words, the noise that affects a particular visual process does not include all frequencies of noise capable of being represented on the display. Figure 3.7 provides an illustration of the eye filter acting on the incoming signal and noise. This figure is provided as an aid to understanding that the RMS noise in Equation 3.7 must be spatially filtered in order to get accurate predictions of CTF_n.

The eye exhibits behavior that seems to imply the presence of selective spatial frequency channels. Exposure to one bar pattern or sinewave grating can affect the visibility of a second pattern. This effect is termed *masking*. Masking only occurs, however, if the bar patterns are close to the same size and oriented in the same direction (Legge, 1987).

The extent to which noise masks a signal depends on the spatial frequency of the signal and the spectral content of the noise (Stromeyer, 1972; van Meeteren, 1988; Greis, 1970). That is, the noise spectral density might not be constant over the frequency limits being considered. If the noise spectral density is not constant, then the noise is “colored.” The ability of colored noise to mask a signal depends on the relative position of the signal and noise in the frequency domain. CTF_n depends on the power spectral density of the noise rather than total noise power (Raghaven, 1989).

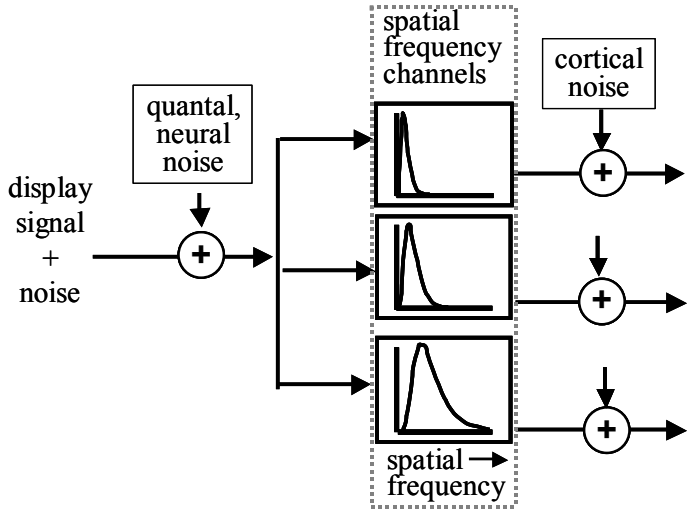
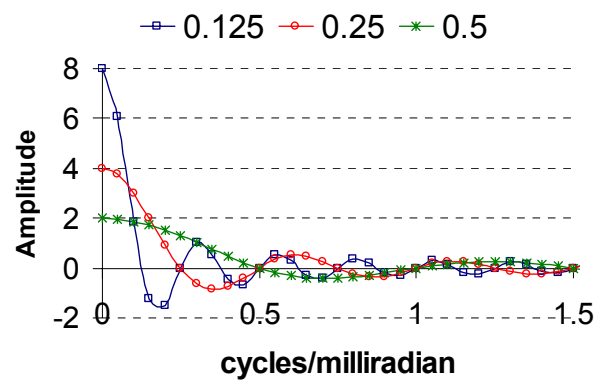


Figure 3.7 Spatial Filter Acts Upon Incoming Signal and Noise

Figure 3.8 shows the visual filters proposed by Barten (1999) based on fit to psychophysical data. The filters shown are for 0.125, 0.25, and 0.5 cycles per milliradian sinusoidal gratings. Equation 3.8 gives the formula for the Barten eye filter B(ξ). ξ_0 is the frequency of the sinewave grating. When using Barten's formulation, the signal is expressed as modulation.

Figure 3.8 Illustration of Three Barten Eye Filters



$$B(\xi) = \exp\left\{-2.2\left[\log\left(\frac{\xi}{\xi_0}\right)\right]^2\right\} \tag{3.8}$$

As verified by numerical integration of Equation 3.8, the bandwidth of Barten's filters increases in proportion to ξ_0. Given a level of white noise, signal to noise increases in proportion to the square root of bar size. This is because, with noise power spectral density constant over the frequency band of interest, the noise associated with a filter is proportional to the square root of bandwidth. So Barten's filters work in lieu of the matched filters for white noise. This has also been verified by using both types of filters with Equation 3.7 to predict the image intensifier experiment reported by Vollmerhausen (1995). Both eye filters give identical results in white noise.

However, the Barten filters also work with colored noise. Barten compares predicted and experimental results from two researchers (van Meeteren, 1988; Stromeyer, 1972). Barten's eye filters also predict the results of Chen (1994).

To illustrate the benefit of the Barten filters over the matched filters, consider the Air Force 3-bar chart shown in Figures 3.9a through c. In Figure 3.9b, low frequency noise is superimposed on the chart patterns. In Figure 3.9c, the image is corrupted with high frequency noise; low spatial frequency noise has been filtered out. The standard deviation of the noise is the same for both Figures 3.9b and 3.9c. Looking at the high frequency bars (the small bars to the right, center of the picture), the high frequency noise masks the bars more than the low frequency noise. If the eye were simply integrating over a bar area, the low frequency noise would actually be more effective in masking the high frequency bars.

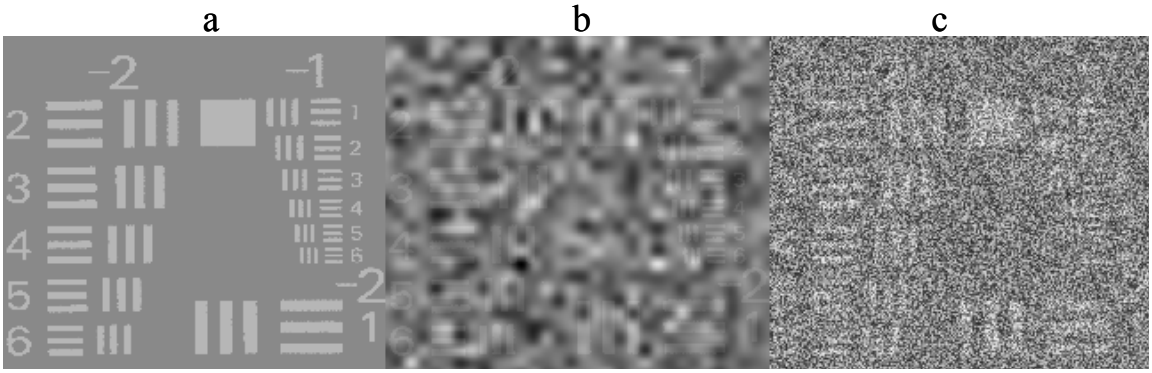
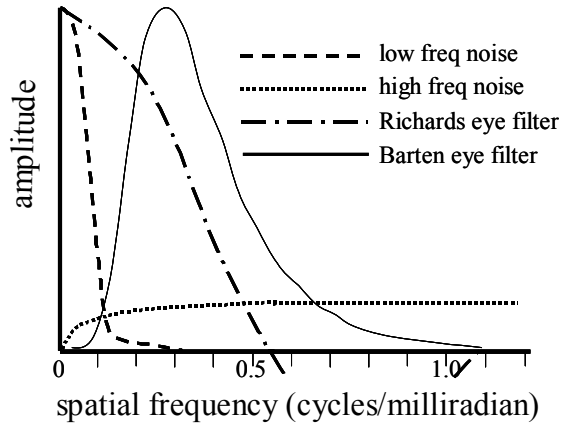


Figure 3.9 Air Force 3-bar Chart Corrupted d by Low (b) and High (c) Frequency Noise

This is illustrated by Figure 3.10. That figure shows the spectra for both the low and high frequency noise. Also shown are both the matched bar filter and the Barten eye filter associated with the smallest 3-bar pattern in Figure 3.9. For this discussion, it is assumed that the bar charts are viewed from a distance five times the width of one chart; Figure 3.9 is three charts wide. The smallest 3-bar pattern is group –1 pattern 6 with a spatial frequency of about 0.275 cycles per milliradian when viewed from a distance five times the width of one chart.

Figure 3.10 Plot of Noise Spectra and Eye Filters The low frequency noise associated with Figure 15b and the high frequency noise associated with Figure 15c are plotted. Also shown are the Richards and Barten eye filters associated with the smallest bar pattern.



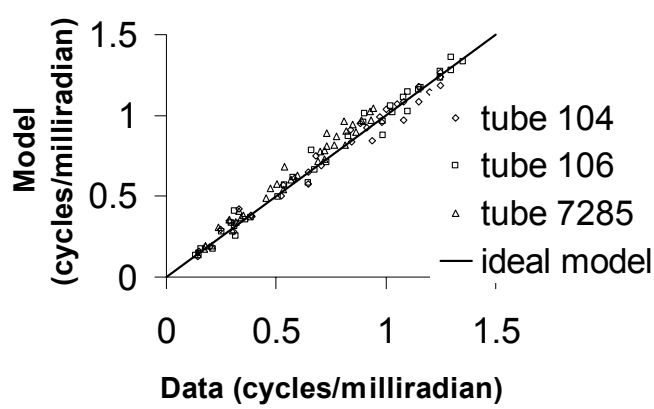
Because the matched filter represents an integration over the bar area, the matched filter has a better response at DC than at higher spatial frequencies. The matched filter cannot explain masking. However, the Barten filters are consistent with observed masking behavior. Equation 3.7 with the Equation 3.9 eye filters provides the foundation for predicting the image quality of imaging sensors.

3.4 Model Calibration

The calibration constant (α) in Equations 4.5 and 4.8 is 169.6 (unitless). This value does not change experiment by experiment or for different sensor types. This value is constant regardless of the system or environment modeled. The value of α was obtained from an image intensifier (I^2) experiment (Vollmerhausen, 1995). During the experiment, Air Force 3-bar charts were viewed through image intensifiers to determine limiting resolution versus chart illumination. The experiment was done with both high contrast (near 1.0) and moderate contrast (0.3) charts. This was an excellent experiment for determining α for several reasons. The physical characteristics of the sensors were accurately measured. The measurements were made at illumination levels from 2.88 E-6 foot candles to 3.39E-3 foot candles. This variation in illumination means that the tubes were operated from noise limited to resolution limited conditions. Measurements were made both with and without laser eyewear protection that reduced the light to the eye by a factor of ten. Also, the tubes used represented both typical and very good MTF, and each tube was operated at three gain levels (25000, 50000, and 75000). Light to the eye varied from as little as 3.6E-4 foot Lamberts (fL) to as much as 12.4 fL. This was an excellent data set because of the controlled nature of the physical sensor data, the wide range of scene illuminations, and the large variation of light to the eye.

Figure 3.11 compares the laboratory data to model results for all 105 data points. The abscissa plots the observed bar resolution and the ordinate is model resolution predictions. If the model were perfect (and if the signal to noise, gain, and MTF measurements of the tube and optics were perfect), then all the points in Figure 3.11 would lie on the straight line. The model predictions are excellent; the square of the Pearson coefficient is 0.98 and the RMS error is 0.057.

Figure 3.11 Plot showing experimental I^2 data versus model predictions. Perfect predictions would lie on the diagonal line.



3.5 Validity of Weber's Law

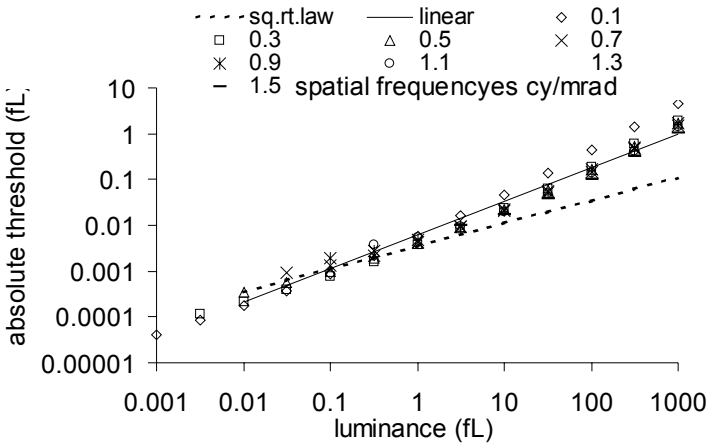
The CTF_n model assumes contrast constancy. That is, threshold luminance increases in proportion to display luminance; this is Weber's Law. Many researchers would object that this is not the behavior of a normal square law detector. Further, certainly a system as highly evolved as the human eye would better approach the theoretical limits represented by the de-Vries-Rose Square Root Law. Perhaps, however, the eye is not a normal square law detector. And perhaps it is not optimized for liminal photon detection.

From a purely physical standpoint, photo-chemicals in the eye are leached out by light; absolute quantum efficiency of the eye decreases as illumination increases. Further, it is certainly not unreasonable to assume that the eye-brain system is optimized for higher order discriminations. Perhaps contrast constancy and color constancy provide the visual system a way of adapting to changing environments. In an evolutionary sense, it may be that a visual system which responds uniformly at sunrise, noon, sunset, in the open, in a cave, or under the shade of a tree is more important than the absolute level at which a faint light is detected on a dark night.

Unfortunately, experimental support for Weber's Law is mixed; some experiments support the idea of contrast constancy over a large variation in illumination, other experiments do not. It is necessary, therefore, to discuss this assumption in more detail as it relates to our CTF_n model.

Figure 3.12 plots absolute threshold versus display luminance for spatial frequencies between 0.1 and 1.5 cycles per milliradian (cy/mrad). The threshold predictions are based on Barten's CTF numerical fit (Barten, 2000) and on a numerical fit to eye MTF (see Appendix E). The figure also shows a straight line; if Weber's Law were exact, then all of the CTF data would lie on a straight line. A dotted line representing the de-Vries-Rose Square Root Law is also shown. The exact ordinate position of the lines is not relevant, because experimental data are used to calibrate the model. The question addressed here is the functional relationship between CTF and luminance.

Figure 3.12 Plot showing CTF at 8 spatial frequencies for luminance levels between 0.001 and 1000 fl. The solid, straight line represents "Weber's Law." The dotted line represents the "Square Root Law."



The CTF data are adjusted to remove the effect of MTF variations which result from differences in pupil dilation as luminance changes. A correction for pupil dilation is

included in the eventual model which is described in Part 4; the pupil-related changes in CTF should not be included in the current discussion.

Because of eyeball MTF and the functioning of the visual cortex, each spatial frequency transitions at a different light level from the Square Root Law region to the Weber's Law region and eventually to saturation. Low frequencies are seen at very low luminance levels. High spatial frequencies require more light. So Weber's Law is applicable to different spatial frequencies at different light levels. This is reflected in the figure by the absence of high spatial frequency predictions for low luminance.

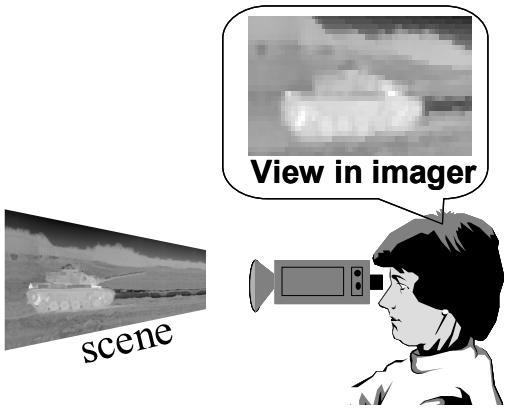
Weber's Law is not exact, but it better fits our needs than the Square Root Law. Looking at Figures 3.2, 3.4, and 3.12, the CTF_n model can be expected to provide accurate predictions for display luminances between 0.01 and 100 fL and approximate predictions from about 0.001 to about 1,000 fL.

4

CONTRAST THRESHOLD FUNCTION OF AN IMAGER

The observer in Figure 4.1 is viewing the scene through an imager and trying to identify the target. The imager helps her by magnifying the target and by permitting the observation of illumination not normally visible to the eye. However, the camera and display add noise and blur. This section describes the observer's Contrast Threshold Function when looking through the imaging system. The system Contrast Threshold Function (CTF_{sys}) is the naked-eye CTF degraded by the amount necessary to account for the blur and noise added by the imager.

Figure 4.1 Observer viewing scene through an imager is trying to identify the target.



As an aid in understanding the formulas for CTF_{sys}, a simple imaging system is illustrated in Figure 4.2. An objective lens focuses light onto a focal plane array (FPA) of detectors. Photo-current is generated over the active area of an individual detector; the active area is indicated by the hatched areas shown in the inset. The scene is blurred because of diffraction and aberrations in the objective lens; the scene is also blurred because of the finite size of the active detector area. The signal is the total photo-current in each detector; shot noise is added to the signal by the statistical nature of the photo-detection process. The individual detector samples are electronically formatted, perhaps filtered

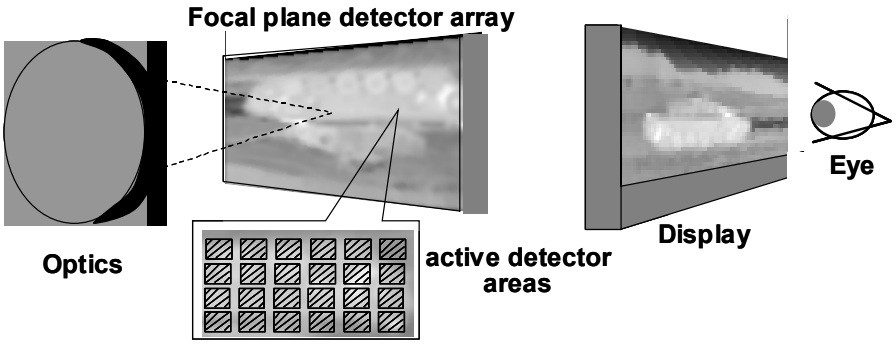


Figure 4.2 Illustration of a staring imager. Optics, detector, display, electronics, and eye all blur the image. Noise is added to the signal during photo-detection.

electronically, and then displayed. So blur can be added by the electronics. The display pixels have a finite size, and this also adds blur to the image. Finally, the eyeball adds blur.

The signal is blurred by every component in the image processing chain; the noise is only blurred by components subsequent to photo-detection. In the model, the Fourier transform of the total, signal blur is called system MTF, whereas the Fourier transform of the blur which filters noise is called the noise filter MTF. Noise filter MTF is a component of system MTF.

4.1 Effect of Blur on Imager CTF

The effect of noise on CTF has been discussed, but the effect of blur has not yet been quantified. In Figure 4.3a, the sinewave chart is just visible to the observer. In 4.3b, an optical system has been introduced between the display and the eye, reducing the visible modulation to below threshold. Assume unity magnification and that the telescope MTF is $H_{sys}(\xi)$. In 4.3c, the displayed modulation has been increased so that the sinewave is once again visible. The display modulation must be increased by the amount lost in the optics. Equation 3.7 for CTF_n is modified as shown in Equation 4.1 to yield CTF_{sys}. CTF_{sys} is the Contrast Threshold Function through the imager; it degrades naked-eye CTF by the amount necessary to account for imager noise and blur.

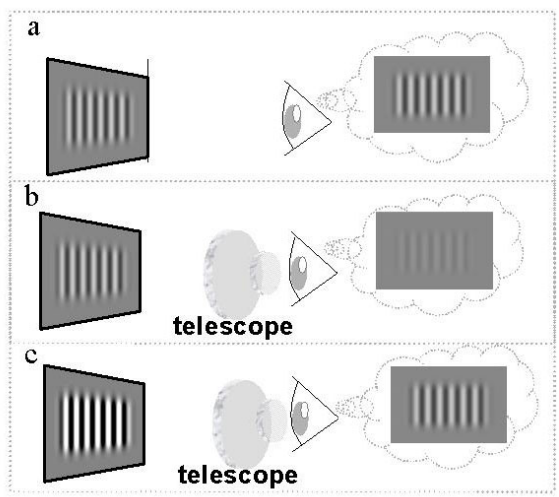


Figure 4.3 Just-visible sinewave modulation in (a) is decreased by the introduction of the telescope in (b). The display modulation must be increased for the sinewave to once again be visible in (c).

$$CTF_{sys}^2(\xi) = \frac{CTF^2(\xi)}{H_{sys}^2(\xi)}\left(1 + \frac{\alpha^2\sigma^2}{L^2}\right) \tag{4.1}$$

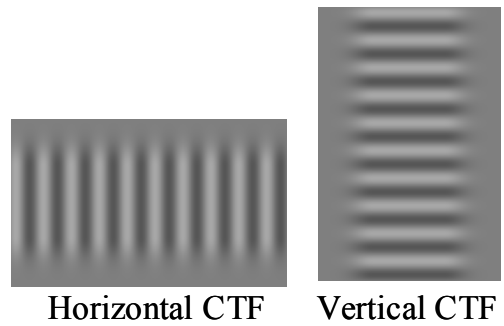
Equation 4.1 is one-dimensional, but imagery has two dimensions. In our models, sensors are analyzed in the vertical and horizontal directions separately, and a summary performance calculated from the separate analyses. The point spread function, *psf*, and the associated MTF are assumed to be separable in Cartesian coordinates. The

separability assumption reduces the analysis to one dimension so that complex calculations that include cross-terms are not required. This approach allows straightforward calculations that quickly determine sensor performance.

The separability assumptions are almost never satisfied, even in the simplest cases. There is generally some calculation error associated with assuming separability. Generally, the errors are small, and the majority of scientists and engineers use the separability approximation. However, care should be taken not to apply the model to circumstances which are obviously not separable; for example, diagonal dither cannot be modeled correctly, nor can diamond shaped detectors.

Since most imagers do not exhibit the same resolution characteristics in the horizontal and vertical directions, the CTF in each direction must be modeled separately. The sinewave pattern at the left in Figure 4.4 is used to generate horizontal CTF modulation;

Figure 4.4 Charts Used to Generate CTF Modulation The left-hand chart is used for horizontal CTF; the right-hand chart is used for vertical CTF.



the pattern to the right is used to generate vertical CTF modulation.

Also, most imagers have a magnification different than unity; the scene is magnified and objects look bigger than without the imager. In our models, the calculations are done in the spatial frequency domain associated with object space. Spatial frequency at the eye (ξ_{eye}) is related to spatial frequency in object space (ξ) by:

$$\xi eye = \xi / SMAG \tag{4.2}$$

where SMAG is the system magnification.

Equations 4.3 and 4.4 give the horizontal and vertical noise bandwidths, respectively, which are associated with calculating horizontal system CTF ($CTFH_{sys}$). The formula for $CTFH_{sys}$ is given in Equation 4.5. Equations 4.6 and 4.7 show the horizontal and vertical noise bandwidths, respectively, for vertical system CTF ($CTFV_{sys}$). Equation 4.8 shows the calculation of $CTFV_{sys}$. In these equations, ξ' and η' are dummy variables with units of cycles per milliradian in object space. The integrations are over all frequencies.

$$QH_{hor}(\xi) = \int \left| B(\xi'/\xi) H_{elec}(\xi') H_{dsp}(\xi') H_{eye}\left(\xi' / SMAG \right) \right|^2 d\xi' \tag{4.3}$$

$$QV_{hor} = \int \left| V_{elec}(\eta) V_{dsp}(\eta) H_{eye}\left(\eta / SMAG \right) \right|^2 d\eta \tag{4.4}$$

$$CTFH_{sys}(\xi) = \frac{CTF\left(\xi/SMAG\right)}{H_{sys}(\xi)}\left(1 + \frac{\alpha^2 \rho\, QH_{hor}(\xi)\, QV_{hor}}{L^2}\right)^{1/2} \tag{4.5}$$

$$QH_{ver} = \int \left| H_{elec}(\xi) H_{dsp}(\xi) H_{eye}\left(\xi/SMAG\right)\right|^2 d\xi \tag{4.6}$$

$$QV_{ver}(\eta) = \int \left| B(\eta'/\eta) V_{elec}(\eta') V_{dsp}(\eta') H_{eye}\left(\eta'/SMAG\right)\right|^2 d\eta' \tag{4.7}$$

$$CTFV_{sys}(\eta) = \frac{CTF\left(\eta/SMAG\right)}{V_{sys}(\eta)}\left(1 + \frac{\alpha^2 \rho\, QH_{ver}\, QV_{ver}(\eta)}{L^2}\right)^{1/2} \tag{4.8}$$

In these equations,

ξ = horizontal spatial frequency in (milliradian)$^{-1}$
η = vertical spatial frequency in (milliradian)$^{-1}$
ρ = detector noise power spectral density in units of (fL-milliradian)^2
L = display luminance
SMAG = angular magnification
B(ξ or η) = the Equation (3.9) eye filters
H_{eye}(ξ or η) = eyeball MTF
H_{elec}(ξ) = horizontal electronics MTF
V_{elec}(η) = vertical electronics MTF
H_{dsp}(ξ) = horizontal display MTF
V_{dsp}(η) = vertical display MTF
H_{sys}(ξ) = horizontal system MTF
V_{sys}(η) = vertical system MTF
QH_{hor} = horizontal noise bandwidth for $CTFH_{sys}$
QV_{hor} = vertical noise bandwidth for $CTFH_{sys}$
QH_{ver} = horizontal noise bandwidth for $CTFV_{sys}$
QV_{ver} = vertical noise bandwidth for $CTFV_{sys}$

4.2 Effect of Contrast Enhancement on Imager CTF

An assumption used in the derivation of Equations 4.1, 4.5, and 4.8 is that the luminance variations on the display are proportional to the luminance or temperature variations in the scene. If the imager has independent gain and level controls, this proportionality can be lost. In fact, since gain enhancement can improve target acquisition performance, it is likely that proportionality will not exist under low contrast conditions.

Contrast enhancement is achieved by gaining the signal and then lowering the display brightness back to the original value. This is illustrated in Figure 4.6. In panel (a), the display luminance is proportional to scene variations in luminance (or temperature for thermal imagers). The figure shows an average luminance (L) and a change in luminance

(ΔL). In panel (b), the display luminance is gained by a factor κ_{con}. All display luminance values increase, including the average display luminance. In panel (c), the display brightness control is used to decrease average display brightness back to the original value (L). However, the change in luminance (ΔL) is now (κ_{con} ΔL). The contrast has increased by κ_{con}.

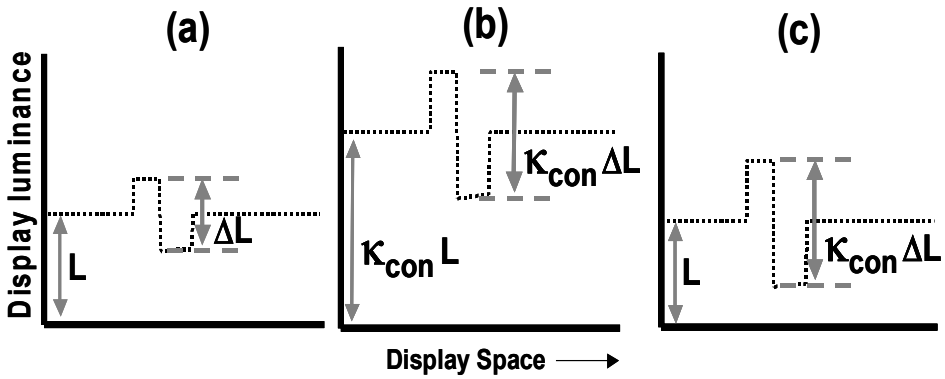


Figure 4.6 Panel (a) shows display luminance proportional to scene variations in luminance or temperature. Panel (b) shows a signal gain of κ_{con}. In (c), average luminance is the same as in (a), and display contrast has increased by κ_{con}.

While gain enhancement does increase perceived noise, noise only increases in proportional to signal. The net effect of gain enhancement is to reduce the impact of eye contrast limitations on performance. With an electronic contrast improvement of κ_{con}, Equation 4.1 becomes:

$$CTF_{sys}^2(\xi) = \frac{CTF^2(\xi)}{H_{sys}^2(\xi)}\left(\frac{1}{\kappa_{con}^2} + \frac{\alpha^2\sigma^2}{L^2}\right). \tag{4.9}$$

Similarly, Equations 4.5 and 4.8 for horizontal and vertical CTF_{sys} become:

$$CTFH_{sys}(\xi) = \frac{CTF\left(\xi/SMAG\right)}{H_{sys}(\xi)}\left(\frac{1}{\kappa_{con}^2} + \frac{\alpha^2\rho QH_{hor}(\xi)QV_{hor}}{L^2}\right)^{1/2} \tag{4.10}$$

$$CTFV_{sys}(\eta) = \frac{CTF\left(\eta/SMAG\right)}{V_{sys}(\xi)}\left(\frac{1}{\kappa_{con}^2} + \frac{\alpha^2\rho QH_{ver}QV_{ver}(\eta)}{L^2}\right)^{1/2}. \tag{4.11}$$

4.3 The Effect of Display Glare on Imager CTF

In Figure 4.7, the soldier's ability to see the display depends on the environment; sunlight reflecting off the display surface can hide the image. Display glare can also be caused by maladjustment of the display brightness control. Whatever the cause, glare can seriously

degrade targeting performance. Glare represents a reduction in contrast at all spatial frequencies. The display modulation loss is:

$$M_{dsp} = \frac{L - L_{glare}}{L + L_{glare}} \tag{4.12}$$

where L_{glare} is the glare luminance and L is the average display luminance. Equations 4.1, 4.5, and 4.8 now become:

$$CTF_{sys}^2(\xi) = \frac{CTF^2(\xi)}{M_{dsp}^2 H_{sys}^2(\xi)} \left(\frac{1}{\kappa_{con}^2} + \frac{\alpha^2 \sigma^2}{L^2} \right). \tag{4.13}$$

$$CTFH_{sys}(\xi) = \frac{CTF\left(\xi / SMAG \right)}{M_{dsp} H_{sys}(\xi)} \left(\frac{1}{\kappa_{con}^2} + \frac{\alpha^2 \rho QH_{hor}(\xi) QV_{hor}}{L^2} \right)^{1/2} \tag{4.14}$$

$$CTFV_{sys}(\eta) = \frac{CTF\left(\eta / SMAG \right)}{M_{dsp} V_{sys}(\xi)} \left(\frac{1}{\kappa_{con}^2} + \frac{\alpha^2 \rho QH_{ver} QV_{ver}(\eta)}{L^2} \right)^{1/2}. \tag{4.15}$$

Equations 4.14 and 4.15 describe quality of vision when using an imager. Different types of electro-optical sensors are modeled by analyzing the blurs and noise associated with the particular technology. Specific formulations for $CTFH_{sys}$ and $CTFV_{sys}$ for various

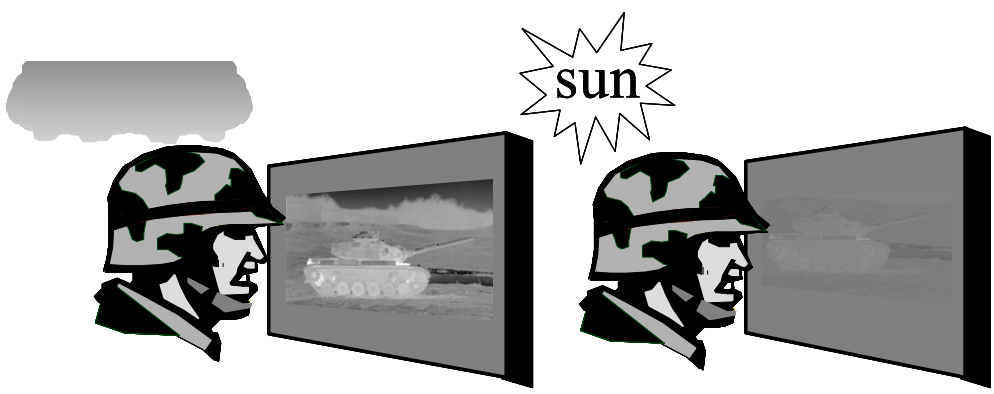


Figure 4.7 At left, clouds are obscuring the sun, and the soldier sees the display clearly. At right, the sun is out, and glare from the display hides the underlying image.

types of imagers are derived later in this report.

4.3 Limit on Modulation Gain

Electronic or digital processing can boost intermediate and high spatial frequencies, improving the displayed representation of scene detail and enhancing target acquisition

performance. An example of high frequency boost is discussed in Appendix B; in that example, the high frequencies of a blurred image are boosted by a factor of eight, and the peak "after boost" image modulation is a factor of 1.7 greater than in the original, un-blurred image. In that particular experiment, boost increased the probability of correctly identifying targets by about 0.2. Because of the TTP metric and the new eye filters, this type of realistic image improvement can now be modeled.

Since the Static Performance Model was first published in 1975, however, there has been a general confusion about how modulation gain is handled in NVL and NVESD models. The confusion can be clarified by describing how sensor system gain is established in the model.

Sensor gain is established by specifying display minimum and average luminance. For reflected-light sensors, this tells the model the delta display luminance that corresponds to the scene illumination and target to background reflectance differences. For thermal imagers, the system gain is established by specifying scene contrast temperature; this is the scene delta temperature that generates the average display luminance. This indirect method of specifying system gain is much simpler for the model user than requiring that actual gain state be input.

The model user could be asked to specify component-by-component absolute gain. This would mean inputting the responsivity of the detector, the actual gain of any automatic-gain-control electronics, and the gain of the display (voltage input to luminance output). A version of the image intensifier CCD (I^2CCD) model used this method of specifying system gain; the method was universally hated by model users. The I^2CCD model used this approach because, at very low illumination levels, the early I^2 cameras could not output sufficient voltage to drive video displays to the desired output luminance; the model had to estimate available output luminance. Using the model, however, required providing information about electronics and display design not normally available to systems analysts.

Modern imagers, including current I^2 CCD cameras, provide sufficient gain that the operational user can set the display luminance as desired. By understanding the operational user's environment and needs, the systems analyst can make a good estimate of the display luminance which will be chosen by the hardware user. That is, an aviator flying without a pilotage aid will keep luminance from instrumentation displays at 0.1 to 0.3 fL in order to maintain dark adaptation; he wants to see outside as well as see his instruments. On the other hand, if the aviator is using a pilotage aid like the Aviator's Night Vision Imaging System (I^2 goggles) or the Pilot's Night Vision System (a thermal imager and helmet display system), then display luminance is typically set in the 1 to 10 fL region; with the higher display luminance, he sees both instrument information and the outside scene better. Generally, the systems analyst can make a reasonable estimate of display luminance if he understands the operational user's task and environment.

In the experiment described in Appendix B, the average display luminance was 5 fL. A display signal modulation of 1.0 at any spatial frequency means that a fully modulated sinewave in the scene would be displayed with a peak-to-peak luminance of 10 fL. Suggesting that the modulation could be 1.7 would mean that the sinewave would have a peak-to-peak displayed luminance of 17 fL and an average luminance of 8.5 fL. This is

not true; the average luminance is 5 fL. A display modulation greater than one makes no physical sense.

So system gain is established by display luminance; it is not established by multiplying the various component MTF. The purpose of component MTF is to establish the relative frequency spectrum of the displayed image. H_{sys} and V_{sys} in Equations 4.14 and 4.15 are normalized to a peak MTF of 1.0.

5

Definition of Target Acquisition Tasks

This section defines target acquisition tasks and discusses model assumptions about accomplishing those tasks. A model is an algorithm or group of inter-related equations based on a set of assumptions. Mathematical models are rigid in their application; they apply only where circumstances match the assumptions. This is certainly true for the target acquisition model.

Two topics are discussed in this section. First, the basic targeting tasks are defined; these tasks include target detection, recognition, and identification (ID). Next, the meaning of the probabilities predicted by the model is described.

The meaning of ***target detection*** varies with operational circumstance. Sometimes a target is detected because it is in a likely place; sometimes a target is detected because it looks like a target (target recognition). Target detection is many things, and therefore not easy to model. Some analysts associate a degree of certainty with target detection; to them, *detection* means the object is of military interest. This is not consistent with current war game modeling; hopefully the war games reflect operational practice.

Search is a process, not a single event, and finding the target generally occurs only after a series of false alarms. The observer searches with the imager in a wide field of view; when an interesting place or object is seen, potentially a target, he switches to a narrower field of view for a closer examination. In our search experiments, with a high density of targets and a good imager, the field of view is typically switched three times before a target is found. With a poorer sensor or a lower density of targets, the field of view is switched many times. When a target is finally confirmed in the narrow field of view, it is credited as a detection in the wide field of view. When analyses are performed to determine the resolution requirements needed to detect the target, the characteristics of the wide field of view are used. The result is that, experimentally, very few "cycles on target" are needed for detection. It must be remembered, however, that the low cycle criteria are associated with a high false alarm rate.

Further, it often occurs that the target and sensor together play a minor role in determining the probability of detection. On the left in Figure 5.1, the target is easily found. This is a thermal image, and the target is much hotter than anything else in the scene. On the right in that figure, the same target is in the same location with the same target to background contrast; only the background objects have changed. The target is hard to find because of clutter. Clutter can affect target acquisition range *by a factor of four.* Very few sensor design parameters have that much influence on range performance.

Search and detection are important sensor functions, and the war-game community pays a great deal of attention to modeling search. However, search modeling is very complex and involves many factors beyond sensor performance. These factors will not be discussed further.

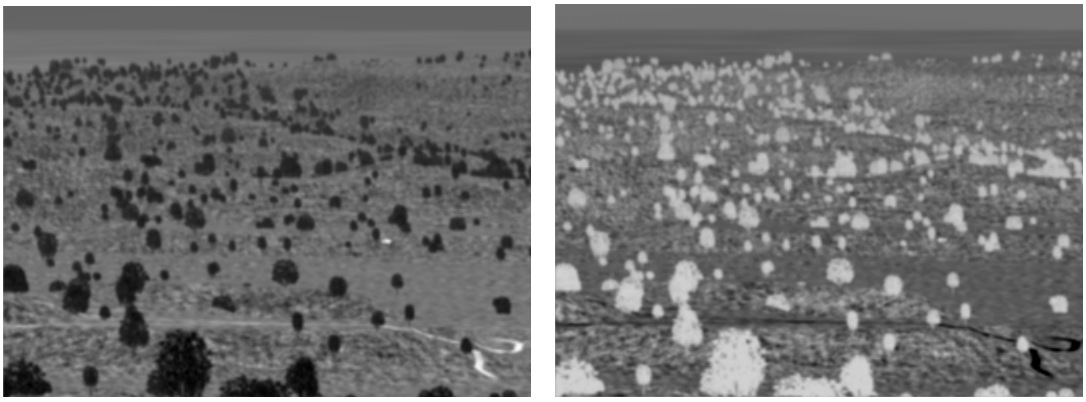


Figure 5.1 On left, hot target in uncluttered background viewed with thermal imager. On right, same target but background has become much hotter.

Recognition involves discriminating which class of vehicle the target belongs in. In Figure 5.2, there are two trucks, two Armored Personnel Carriers (APC) and two tanks. In a recognition experiment, the observers (subjects) are trained and tested on the specific target set. These trained observers are shown the targets at range (so the images are blurred, noisy, perhaps poorly sampled) and asked to specify tank, truck, or APC. If the observer gets the class correct, the task is scored as correct. That is, the observer might mistake the T72 tank in Figure 5.2 for the Sheridan tank. He has correctly "recognized" that the target is a tank. It does not matter that he incorrectly identified the vehicle.

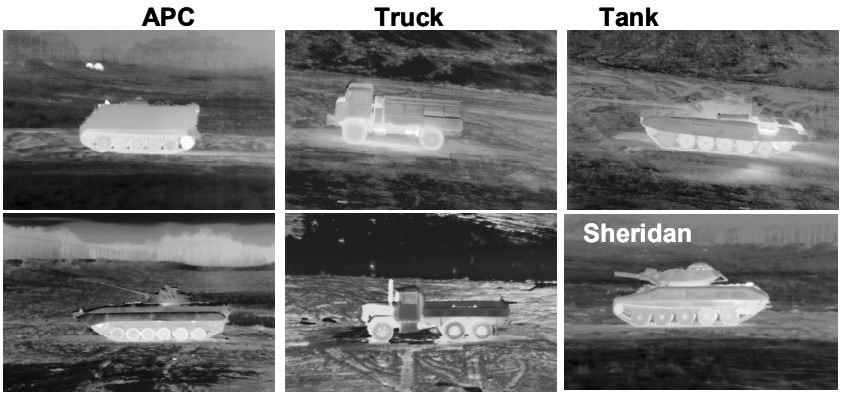


Figure 5.2 Side views of a group of vehicles that might be used in a recognition test.

Note two important things about recognition. First, the difficulty of recognizing a vehicle depends on the vehicle itself and *on the alternatives or confusers.* Discriminations are always comparisons. Task difficulty is established by the set, not by an individual member of the set. Second, a recognition set of targets like the one shown in Figure 5.2 involves easy discriminations and more difficult discriminations. APCs look much more like tanks than either tanks or APCs look like trucks. So the typical recognition task is actually a combination of easy discriminations and more difficult discriminations with

the results averaged. In terms of range performance, the tasks should be modeled separately.

Target identification requires the observer to make the correct vehicle choice. A set of targets which might be used in an ID experiment is shown on the left in Figure 5.3. Only one aspect of each vehicle is shown; experiments use several aspects of each vehicle. Twelve aspects of the T62 Russian tank are shown to the right in Figure 5.3. Again, the observers are well trained and tested that they can correctly identify each individual vehicle. The targets are put at range (blurred, noisy, perhaps corrupted by poor sampling) and the observer must indicate which target he is shown. In this case, the observer must correctly identify the target, not just the class. Calling a T72 tank a T62 tank is scored as an incorrect choice.

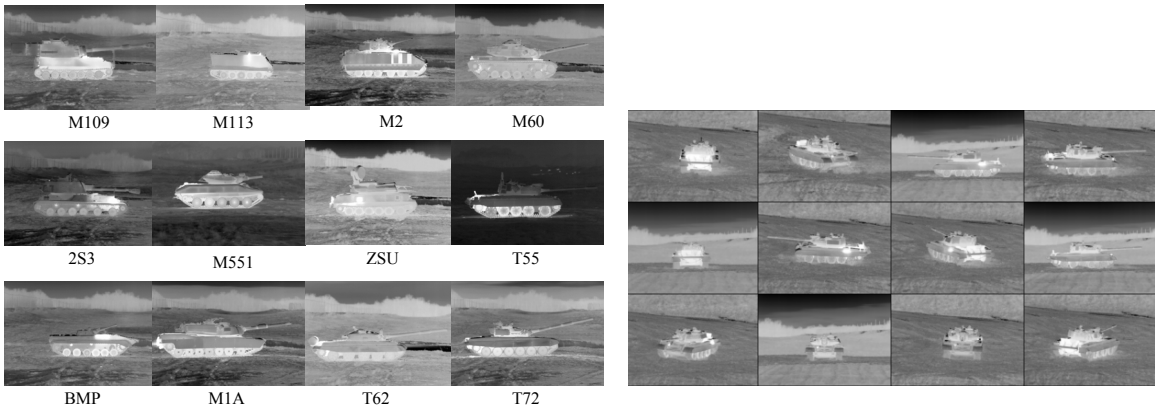


Figure 5.3 Target set for ID experiments.

The difficulty of the ID task depends on the group of targets selected, not the individual target which happens to be within the sensor FOV. The model does not predict the probability of identifying or recognizing individual vehicles; the model predicts the average probability of correctly identifying all members of the group at range. The difficulty of the task depends on how much the members of the group look alike. In Figure 5.4, three observers are trying to identify three vehicles. If the first observer gets all three vehicles correct, the second observer gets two correct, and the third observer gets one correct, then the probability of correct ID is 0.67. That is, six total correct calls divided by nine total calls. This average over both observers and targets in the group is what the model predicts.

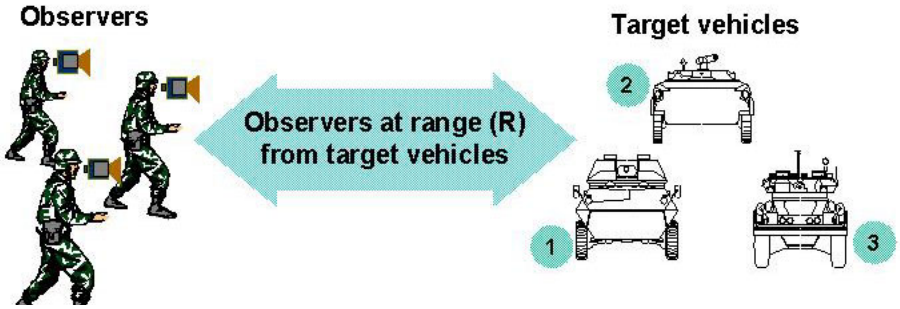


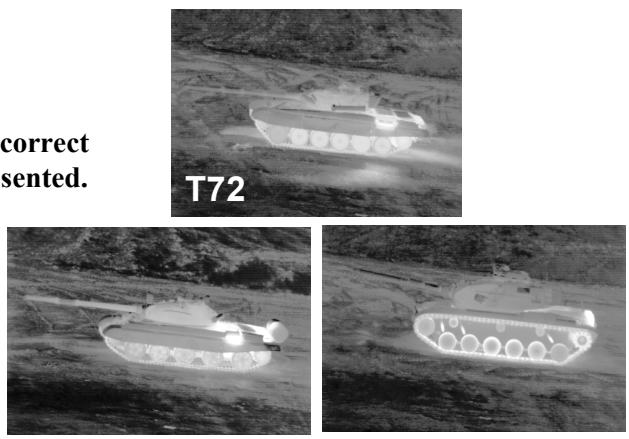
Figure 5.4 Illustration of How Probabilities are Calculated

To achieve prediction accuracy, the model requires a group of observers (ten to twenty) and a group of "like" targets. The group of vehicles in Figure 5.3 are sufficiently alike that model accuracy is good (less than 0.05 average error in the predicted probabilities, with the biggest errors occurring at the 0.5 point in the curve where statistical variability is expected).

A target acquisition discrimination is always a comparison. Is it target A or target B or target C? Is it a target or background? It is quite common for an analyst to be asked the question: "Using this sensor, at what range can I identify a T72 Russian Tank?" That question cannot be answered; it is only partly formulated. The examples below might clarify this statement.

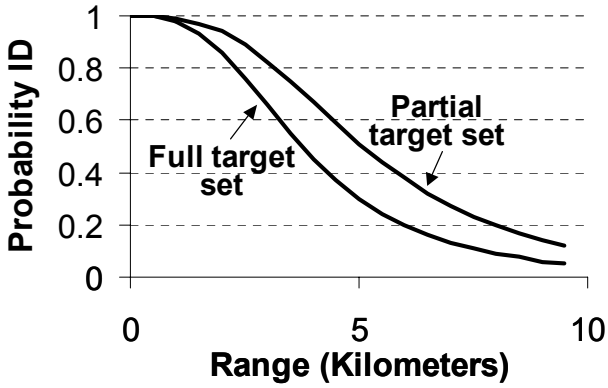
The Iraqi's used T72 Tanks in the 1991 war. One U.S. ally in that war was Egypt; because of the vagaries of the cold war era, Egypt owns both T62 Russian Tanks and U.S. built M60 Tanks. The three tanks are shown in Figure 5.5. Because Russian tanks tend to look alike, if our ally used a T62 Tank, the friend-versus-foe decision would be more difficult than if our ally used an M60 Tank. So the range at which a T72 can be reliably identified depends on the alternative.

Figure 5.5 Images of three tanks illustrating that the probability of correct ID depends on the alternatives presented.



ID experiments have been performed using the target set shown in Figure 5.3. Probability of ID versus range is shown in Figure 5.6 by the curve labeled "full target set." The curve labeled "partial target set" shows the results of an ID experiment using nine of the twelve targets; the M109, T62, and T55 have been removed from the target set. The T62 and T55 look like the T72; the M109 looks a lot like the 2S3. Removing these vehicles makes identifying the remaining targets easier, and this results in a higher probability of ID.

Figure 5.6 Probability of ID versus range for the full set of target shown in Figure 5.3 and for an easier to identify, partial set where the T62, T55 and M109 vehicles are not used.



Target detection, recognition, or identification is determined by a process of seeing viewpoint-invariant details. The size, contrast, and number of characteristic details visible to the observer determines the probability of target acquisition. Our model predicts the quality of the image and therefore the ability of the observer to acquire the target. However, targets are acquired by differentiating them from the possible alternatives. This means that the features which uniquely define a target are those which differentiate that target from other targets or from background. Therefore, task difficulty depends on how alike the targets look or the level of target-like clutter in the background.

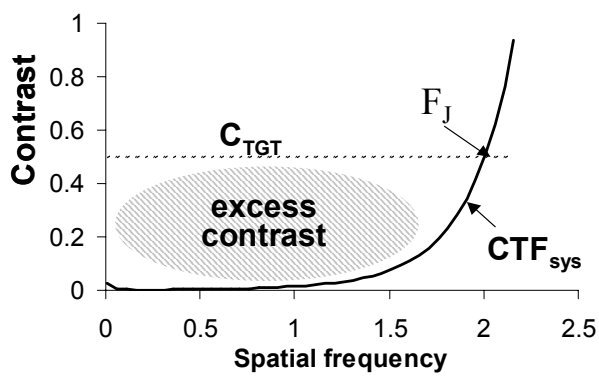
6

PREDICTING TARGET ACQUISITION PERFORMANCE FROM IMAGER CTF

Both the Johnson criteria and the Target Task Performance (TTP) metric are MTF-based metrics. These metrics share the concept that image quality can be quantified by a weighted integral over spatial frequency of the ratio between signal and CTF. It is assumed that the excess modulation over threshold provides the information acted upon by the visual system. A great virtue of MTF-based metrics is the simplicity of implementing a range performance model; for a specific task, it is assumed that range is proportional to the metric value.

The Johnson criteria uses the limiting frequency visible at the average target contrast to quantify image quality and therefore range performance. The Johnson metric is defined by the spatial frequency range (F_J) over which the apparent target contrast ($\mathbf{C}_{TGT}$) exceeds the system contrast threshold [$CTF_{sys}(\xi)$]. See Figure 6.1 for an illustration of the Johnson metric.

Figure 6.1 Johnson criteria uses intersect of target apparent contrast and CTF_{sys} as measure of image quality for targeting purposes. In this figure, the intersection occurs at frequency F_J.



The TTP metric gives weight to the amount that threshold is exceeded at each spatial frequency; this makes the TTP metric sensitive to image qualities not quantified by the Johnson methodology. The TTP metric is calculated as shown in Equation 6.1. In this equation, ξ_{cut} is the high spatial frequency where CTF_{sys} exceeds C_{TGT}; ξ_{cut} equals F_J. ξ_{low} is the spatial frequency below which CTF_{sys} exceeds C_{TGT}. Lateral inhibition in the eye results in CTF_{sys} having a spatial bandpass response; the eye sees intermediate spatial frequencies better than either very low or high frequencies. However, ξ_{low} is very nearly zero. Because of the square root, contrast that is well in excess of threshold is not as important as contrast that just exceeds threshold. The TTP value calculated using Equation 6.1 is used in lieu of F_J to quantify image quality and predict range performance.

$$TTP = \int_{\xi_{low}}^{\xi_{cut}} \left[\frac{C_{TGT}}{CTF_{sys}(\xi)} \right]^{1/2} d\xi \qquad \textbf{(6.1)}$$

While the Johnson criteria provides reasonable performance estimates in many circumstances, applying that criteria to sampled imagers generally results in pessimistic predictions. In recent years, modelers have developed "work arounds" to use the Johnson criteria with sampled imagers (Driggers, 2000; Wittenstein, 1999; Bijl, 1998). These fixes have limited application, however, because they are empirical adjustments of a basically flawed model. The Johnson criteria "work arounds" do not permit the modeling of digital image enhancement, for example, because variations in CTF_{sys} below the cutoff frequency do not affect the metric value. The TTP metric does an excellent job of predicting the performance of both well-sampled and under-sampled imagers. It also predicts the performance impact of frequency boost, colored noise, and other characteristic features found in modern imagers. A summary of some of the experimental data supporting the TTP metric and illustrating the problems with the Johnson criteria is provided in Appendices A, B, and C.

6.1 Predicting Probability versus Range

A range performance model is created by assuming that target acquisition range is proportional to the image quality metric. That is, the range at which a task can be performed is proportional to the TTP value calculated in Equation 6.2. For a given target contrast and size, a given task like target ID, and a selected probability of accomplishing the task, the range is calculated as shown in Equation 6.2.

$$Range = \frac{\sqrt{A_{TGT}}\, TTP}{N_{required}} \qquad (6.2)$$

For tactical vehicle targets, size is usually taken as the square root of the viewed target area (A_{TGT}). $N_{required}$ represents task difficulty and desired probability of success; the value of $N_{required}$ is established experimentally for a particular target set and task. For vehicle images, the zero range target to background contrast is defined by:

$$C_{TGT-0} = \frac{\sqrt{(\Delta\mu)^2 + \sigma_{tgt}^2}}{2\,\mu_{scene}} \qquad (6.3)$$

where μ_{scene} is the average scene luminance (or temperature) in the vicinity of the target, $\Delta\mu$ is the difference in average luminance (temperature) between the target and local background, and σ_{tgt} is the standard deviation of the target luminance (temperature).

While range is proportional to image quality, the probability of accomplishing a task is not. To calculate probability with the target at a given range, first use Beer's law or MODTRAN to calculate the atmospheric transmission (τ), then calculate the apparent target contrast at the sensor (C_{TGT}).

$$C_{TGT} = \tau\, C_{TGT-0} \qquad (6.4)$$

C_{TGT} is found using Equation 6.4, the TTP value is calculated using Equation 6.1, and then resolved cycles is calculated using Equation 6.5.

$$N_{resolved} = \frac{\sqrt{A_{TGT}}\, TTP}{Range} \tag{6.5}$$

An empirically derived Target Transfer Probability Function (TTPF) is used to relate probability of task performance to the ratio of $N_{resolved}$ to V50, where V50 is the metric value needed to accomplish the task with a 0.5 probability. Again, V50 is established experimentally. The TTPF curve is a logistics function as defined by Equation 6.6.

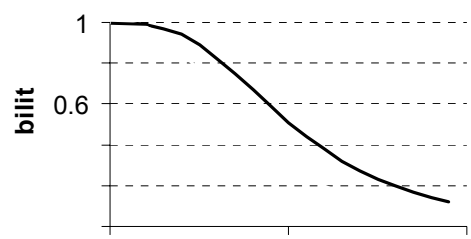
$$P = \frac{\left(N_{resolved} / V_{50} \right)^E}{1 + \left(N_{resolved} / V_{50} \right)^E} \tag{6.6}$$

where

$$E = 1.51 + 0.24\, N_{resolved} / V50 \tag{6.7}$$

The process is repeated at range intervals to generate a probability versus range function as shown in Figure 6.2. If the goal is to predict the outcome of a field experiment, then the probabilities generated with Equation 6.6 are corrected to add chance and to add the 0.1 probability associated with observer mistakes; the probability corrections are described in Section 6.2.

Figure 6.2 Typical model output is probability versus range as shown in the figure.



Target area in Equation 6.5 and target contrast in Equation 6.1 refer to averages over the group of targets involved in the experiment or scenario. The reasoning behind this is discussed in Section 6.3.

Many imagers have different resolution characteristics in the horizontal and vertical dimensions. In scanning thermal imagers, for example, the horizontal resolution is often much better than the vertical resolution. As discussed in Section 4, CTFsys is calculated for the two dimensions using Equations 4.5 and 4.8. Then the TTP metric is calculated separately for each direction.

$$TTPH = \int_{\xi_{low}}^{\xi_{cut}} \left[\frac{C_{TGT}}{CTFH_{sys}(\xi)} \right]^{1/2} d\xi \tag{6.8}$$

$$TTPV = \int_{\eta_{low}}^{\eta_{cut}} \left[\frac{C_{TGT}}{CTFV_{sys}(\eta)} \right]^{1/2} d\eta \tag{6.9}$$

The TTP value to use in Equation 6.5 to find $N_{resolved}$ at each range is then the geometric mean of the horizontal and vertical TTP values.

$$TTP = \sqrt{TTPH\ TTPV}$$

This mean value of TTP is used in Equation 6.6 to find probability of target acquisition.

6.2 Meaning of Model Probabilities

The probabilities predicted by the model are intended to be used to assess sensor "goodness" for target acquisition. Model probabilities have been adjusted to remove the influence of factors which affect target acquisition probability but which are independent of sensor design. The model probabilities have been corrected for chance and corrected for non-ideal observer performance. The relationship between model probabilities and observed data is explained in this section.

If an ID experiment is conducted using four vehicles, then there is a 0.25 probability of correct ID just by chance. As range increases, the measured probability drops to 0.25, not to zero. If twelve targets are used in the experiment, then the probability drops to 0.083 at long range. If a recognition experiment is performed using three classes of targets (tank, truck, and APC), then the probability of getting the answer correct just by chance is 0.33. In a wheeled-versus-tracked classification experiment, probability of correct choice by chance is 0.5 because there are only two choices.

The probability of chance is removed before using experimental data to calibrate the model.

$$\text{Model Probability} = \frac{\text{Measured Probability} - P_{chance}}{1 - P_{chance}} \tag{6.11}$$

Where P_{chance} is the probability of correctly identifying the target or target class just by chance. If four targets or target classes are used in the experiment, then Pchance is 0.25. If twelve targets or target classes are used, then Pchance is 0.083. To compare model predictions with field data, the above formula is inverted.

$$\text{Predicted Measured Probability} = \text{Model Probability}\,(1 - P_{chance}) + P_{chance} \tag{6.12}$$

Another correction is made to experimental data before comparing it to model predictions. Even well trained, conscientious people make mistakes. We observe a 0.1 error rate which cannot be correlated to image quality, training, or apparent motivation. This error rate is fairly consistent across the various target acquisition tasks (search, recognition, identification). Some observers do achieve 1.0 probabilities on clean target image sets, but when an average over twenty observers is made, the top probability is 0.9. Our data asymptotes to 0.9 probability at close range.

Whether this error rate is observed under field conditions is not known by the authors. Whether that error rate should be represented in the model is a matter of judgment. Traditionally, however, this drop in probability due to mistakes has not been included in performance models.

If it is desired to include the base mistake rate for an ensemble of observers, then use Equation 6.13 rather than Equation 6.12 to relate field measured probabilities to model probabilities.

$$\text{Predicted Measured Probability} = \text{Model Probability}\,(0.9 - P_{chance}) + P_{chance} \qquad (6.13)$$

6.3 Field Test Example

In a hypothetical field test, eight tactical vehicles are available: M1, BMP, T72, M109, M113, M2, 2 $^1/_2$ ton truck, and a HMMWV. Since six are tracked vehicles, one is a truck, and the other a HMMWV, the decision is made to drop the truck and HMMWV as being too dissimilar from the rest of the vehicles. The average dimension (square root of area) and average contrast for the six tracked vehicles are 3 meters and 4° C, respectively. A V50 of 20 for identifying this particular group of vehicles is established by experience and expert judgment.

The model is run to predict probability versus range for the sensor system being evaluated. Five ranges are selected which span ID probabilities from high to very low. Because of the vagaries introduced by mistakes, chance, and the many factors which bias real field data, a system should not be evaluated using only close range, high probability data.

At each range, three aspects of all targets are presented. If the vehicle has a rear mounted engine, then the aspects are front, side, and opposite side-rear oblique. If the vehicle has a front mounted engine, then the aspects are rear, side, and opposite side-front oblique. The total test consists of 18 target views at five ranges for a total of 90 images.

Images are collected for viewing in the lab, or observers are taken to the field. Certainly, data taking is simplified if the observers are not in the field. The observer must be deprived of any clues other than the sensor imagery which might help him identify the target. The observer's situational awareness is best limited by separating him from the test site. However, the output of some sensors is not easily recorded for later display, and the experiment is best performed in the field.

The observers are trained to ID the vehicles used in the experiment. The observers must pass a test to prove they can correctly ID all the vehicles before participating in the experiment. The observers are asked to ID the targets from the sensor imagery. The average correct ID probability for each range is calculated based on observer responses. The total experiment yields five (5) data points. To compare the model probabilities to the actual data collected in the field, model probabilities are adjusted using Equation 6.12 or 6.13 above with 0.167 substituted for P_{chance}. The choice of which equation to use depends on the number, experience, and motivation of the observers. Use Equation 6.12 if the experiment involved a few, highly experienced observers. Otherwise, use Equation 6.13.

If the above steps are followed, the model accurately predicts observer performance.

6.4 Estimating Task Difficulty (V50)

There is currently no objective way to establish V50 for a target group other than by experiment. We have found, however, that a careful process of comparative judgment can provide good estimates for V50. That is, knowing the experimentally established value of V50 for example target groupings, comparative judgment can be used to estimate the V50 for a related group of targets.

This sections provides some example target sets and the associated V50 values. Examples are given for detection, recognition, and ID. Since V50 values are based on experience, historical data should also be useful in establishing values to be used in the new model. However, there are several issues to consider when making comparisons between new V50 and old N50 values. In addition to giving V50 examples, this section discusses the differences between historical values of N50 used with the Johnson criteria and values of V50 used with the new TTP metric.

The Johnson metric can be thought of as an integral over spatial frequency.

$$F_J = \int_0^{F_J} [1] d\xi$$

where F_J is the frequency where C_{TGT} equals CTF_{sys}; see Figure 6.1. The "1" in the integral is to emphasize that each frequency increment counts equally; if the target apparent contrast exceeds the threshold needed for visibility at a particular frequency, then that frequency increment is counted in the Johnson bandwidth.

The TTP metric value is also an integral over essentially the same frequency range. The value of ξ_{low} in Equation 6.1 is always small; to a good approximation, ξ_{low} is zero. Remembering that ξ_{cut} equals F_J:

$$TTP = \int_0^{F_J} \left[\frac{C_{TGT}}{CTF_{sys}(\xi)} \right]^{1/2} d\xi \ . \tag{6.15}$$

The ratio C_{TGT}/CTF_{sys} is always greater than one. This means that the value of TTP is always greater than F_J. The ratio between the Johnson metric and the TTP metric is not fixed; if the ratio were fixed, then the two metrics would provide identical performance predictions. However, for those cases where both metrics predict performance well, the ratio of TTP value to Johnson metric value is approximately 2.7:1.

This does not mean that the Johnson N50 values can be multiplied by 2.7 to obtain an V50 for the new model. In the new model, V50 represents the resolved cycles needed to achieve a 0.5 probability independent of chance. Historically, the data used to establish N50 were not corrected for chance.

It is not clear how N50 values for two-class discriminations were established. Although the historical value of N50 for discriminating wheeled vehicles from tracked is 1 to 2 cycles, a 0.5 probability of success is actually achievable with zero cycles. Since there are

only two classes, success half of the time is guaranteed. For 3-class recognition (tank-truck-APC), most of the 0.5 probability is attributable to the 0.33 probability of being correct just by chance. As the number of choices increases, the impact of using uncorrected data to establish N50 decreases.

Table 6.1 shows how an N50 based on uncorrected data must be increased to be used in a model which does remove probability due to chance. This table is based on the TTPF associated with the Johnson metric. The multiplier values are the ratio of N50 needed to achieve 0.5 probability without chance to the N50 needed when chance is included. For example, if a 3-choice recognition experiment (tank-truck-APC) yields an N50 of 3 based on uncorrected data, then the N50 for corrected data would be 3 * 1.79 or 5.37. It is easier to achieve 0.5 probability when chance is included, so the N50 for uncorrected data is smaller than the N50 for corrected data. As the number of choices increases, the impact of chance on the data decreases, and the ratio of the N50 values approaches one.

Table 6.1

Number of choices	3	4	5	6	8	10	12	20
N50 Multiplier	1.79	1.43	1.3	1.23	1.16	1.12	1.1	1.05

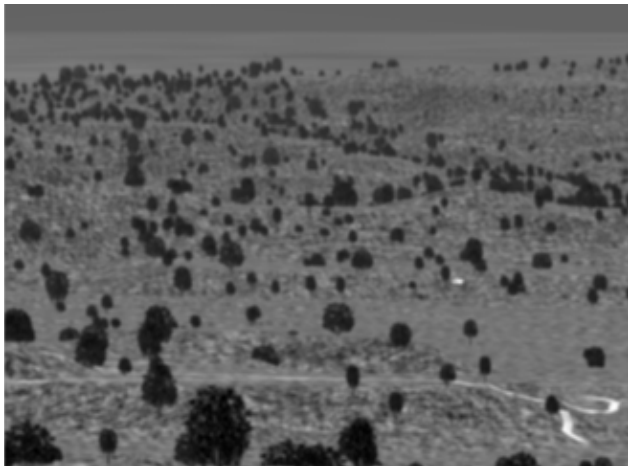
Two examples will illustrate how V50 values for the new model can be derived from N50 values used with the Johnson model. With the Johnson metric, tank-truck-APC recognition is modeled using an N50 of 3. The N50 for data with chance removed is 5.37. Multiplying by 2.7, the value of 14.5 is the V50 for use in the new model.

Although details are not available on how the "standard" N50 of 6 for ID was established, assume that a 6-choice experiment was used. A equivalent V50 for the new model is found by multiplying 6 by 1.23 and then by 2.7 yielding 19.9. Table 6.2 gives Johnson N50 and TTP V50 values for a selection of target acquisition tasks.

Table 6.2

Task description	**N50 w/ chance**	**TTP V50 w chance**	**TTP V50 w/o chance**
Low clutter thermal detect; Figure 6.3	0.75	2	2
Medium clutter thermal detect; Figure 6.4	1.7	4.6	4.6
Recognize tank-truck-APC; Figure 6.5	3	8.1	14.5
Recognize truck/wheeled-armored/tracked-armored Figure 6.6 (Reference Devitt, 2001)	3.5	9.45	16.9
ID 12-target set; Figure 6.7	7.8	21.2	23.3
ID 9-target set; Figure 6.8	6.5	17.6	20.0

Figure 6.3 Example of Low Clutter, Thermal Detect



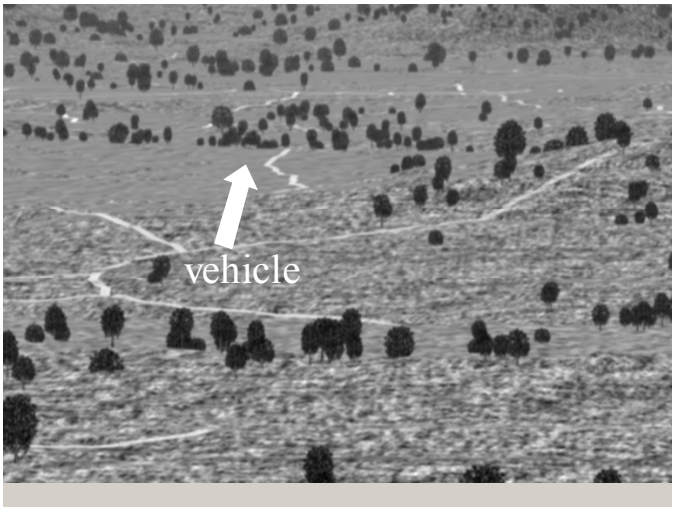


Figure 6.4 Example of Moderate Clutter, Thermal Detect

Figure 6.5 Recognition Tank-Truck-APC Several aspects of each vehicle would be used in a recognition experiment.

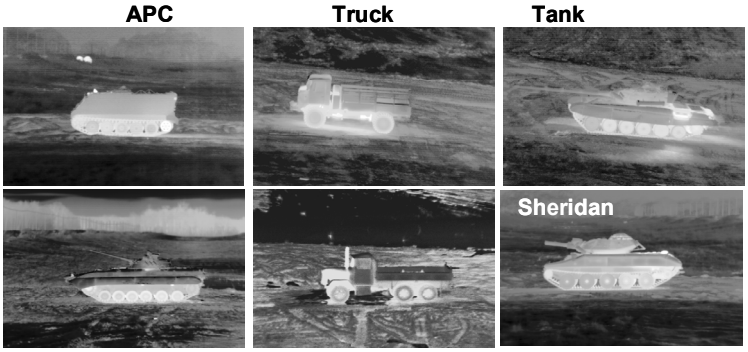


Figure 6.6 Recognition Tracked-armored/Wheeled-armored/Soft-truck Experiment involved many vehicles and aspects; these are examples.

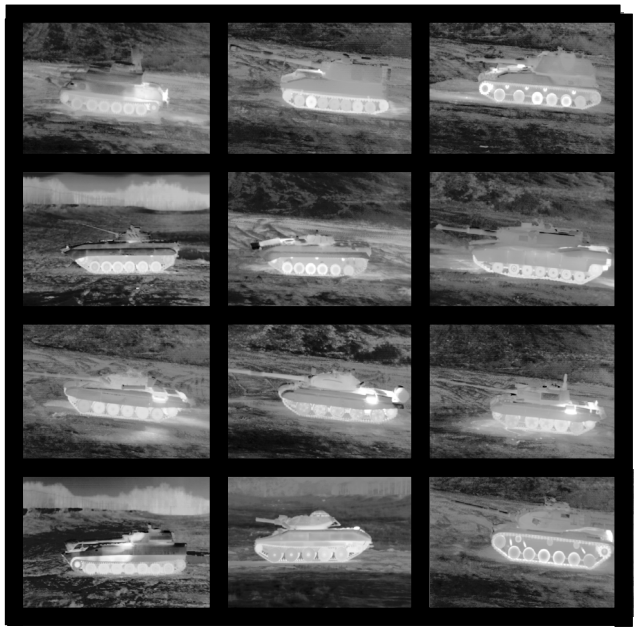


Figure 6.7 Twelve Tracked Military Vehicles

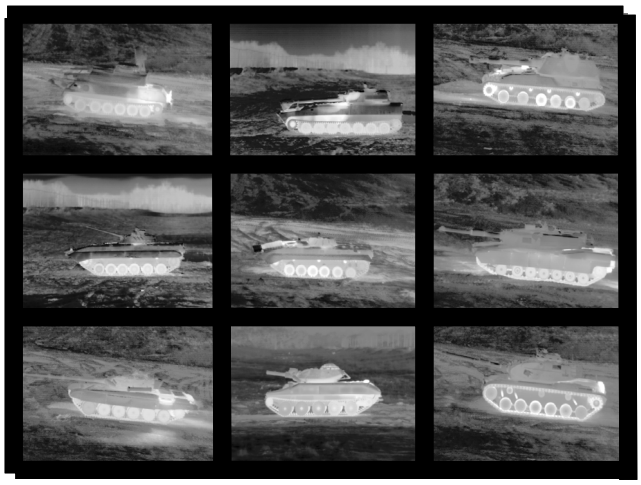


Figure 6.8 Nine Tracked Military Vehicles Three of the vehicles in Figure 6.7 have been removed; since those vehicles look like some of the remaining vehicles, this target set is easier to ID.

7

Modeling Sampled Imagers

The sampling limitations associated with focal plane array (FPA) imagers cause an aliased signal that corrupts the image. Aliasing can cause distortion of scene detail; for example, fence posts can be fatter, thinner, or disappear completely. Aliasing can also cause display artifacts like line raster. The aliased signal is a function of input image, pre-sample blur, sampling frequency, and image reconstruction at the display. The model used to predict the amount of sampling artifacts present in an imager is described by Vollmerhausen (2000, 04/2000).

Aliasing can degrade target acquisition performance. Experiments to calibrate the decrease in performance based on the aliased signal present are described in several references (Vollmerhausen, 1999; Krapels, 1999, Krapels, 2001; Devitt, 1999). The technique for predicting sampling artifacts and the resulting degradation in range performance is summarized here. Examples showing the predictive accuracy of the technique are described in Appendices A and C.

It has become common practice among engineers to use the term *aliasing* to refer only to spurious frequency content that overlaps and corrupts the signal in the original (pre-sampled) frequency band. Sampling actually causes aliasing at all spatial frequencies. However, to avoid confusion about the meaning of aliasing, the term *spurious response* is used in this paper. The part of the image spectrum which results from sampling, other than the original frequency content, is referred to as spurious response. That is, in frequency space, spurious response is the Fourier transform of the sampling artifacts.

The spurious response of a sensor corresponds to artifacts in the sensor imagery; it is a much better indicator of sampling efficacy than the half sample rate. The spurious response of a sensor can be described in a manner very similar to the sensor Modulation Transfer Function (MTF) in that, the frequency components of the spurious response may be plotted similar to an MTF. The greatest barrier in the use of spurious response to characterize sensor performance is the calibration of human reaction to spurious response.

The amount of spurious response in an image is dependent on the spatial frequencies that comprise the scene and on the blur and sampling characteristics of the sensor. However, the spurious response capacity of an imager can be determined by characterizing the imager response to a point source. This characterization is identical to the MTF approach for continuous systems. MTF is a trusted indicator of optical quality. But the need for good MTF cannot be established until the scenario and task are defined. Good MTF is not always needed; it is prized because of the potential is provides. The same is true for the

spurious response characteristics of an imager. The actual amount of aliasing cannot be known without specifying the scene, but the tendency of an imager to generate sampling artifacts is significant in the same sense that good MTF is significant.

The effect of sampling on target acquisition is modeled with the following procedure. First, the spurious response of the imager is analyzed; this is done by characterizing the shift-variant response of the imager to a point source. Once the amount and nature of the spurious response is known, experience from target acquisition experiments with sampled imagery is used to establish the expected drop in performance.

7.1 Response Function of a Sampled Imager

The response function $R_{sp}(\xi)$ for a sampled imager is found by examining the impulse response of the system. This procedure is identical to that used with non-sampled systems. The function being sampled is $h_{pre}(x)$, the point spread function of the pre-sampled image. Assume the following definitions:

ξ = spatial frequency (cycles per milliradian)
ν = sample frequency (samples per milliradian)
d = spatial offset of origin from a sample point (in milliradians)
$H_{pre}(\xi)$ is the pre-sample MTF (optics and detector)
$P_{ix}(\xi)$ is the display MTF (crt spot, sample and hold, eyeball MTF)

Then the response function $R_{sp}(\xi)$ is given by the following equation.

$$R_{sp}(\xi) = \sum_{n=-\infty}^{n=\infty} H_{pre}(\xi - n\nu) e^{-i(\xi - n\nu)d} P_{ix}(\xi)$$

$$R_{sp}(\xi) = H_{pre}(\xi) e^{-i\xi d} P_{ix}(\xi) + \sum_{n \neq 0} H_{pre}(\xi - n\nu) e^{-i(\xi - n\nu)d} P_{ix}(\xi) \tag{7.1}$$

The response function has two parts, a transfer term and spurious response term. The $n=0$ term in Equation 7.1 is the transfer response (or baseband response) of the imager. The transfer response does not depend on sample spacing, and it is essentially the only term that remains for very small sample spacing. A very well sampled imager has the same transfer response as a non-sampled imager.

However, a sampled imager always has the additional response terms (the $n \neq 0$ terms). These terms mathematically describe the spurious response. The spurious response terms in Equation 7.1 are filtered by the display MTF, $P_{ix}(\xi)$, in the same way that the transfer response is filtered. However, the position of the spurious response terms on the frequency axis depends on the sample spacing. Also, the phase relationship between the transfer response and the spurious response depends on the sample phase. See Figure 7.1 for a graphical illustration of the transfer and spurious response terms.

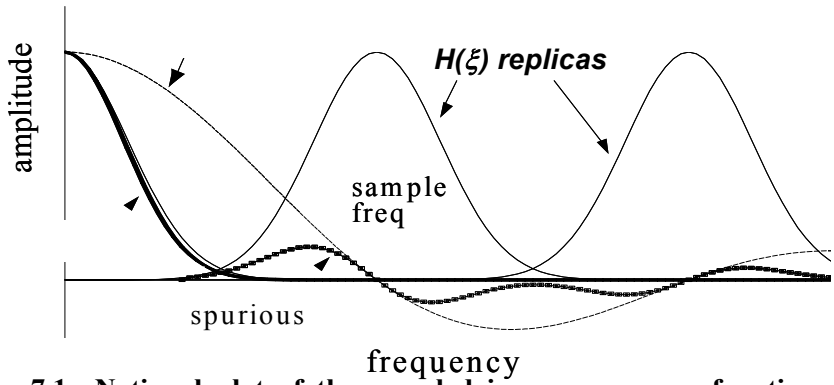


Figure 7.1 Notional plot of the sampled imager response function. The pre-sample MTF *H(ξ)* is replicated at multiples of the sample frequency. The transfer response is the pre-sample MTF multiplied by the display and eye MTF *Pix*(ξ). The spurious response is the pre-sample replicas filtered by *Pix*(ξ).

7.2 Impact of Sampling on Range Performance

A number of experiments have been performed to discover the impact of spurious response on targeting performance. Based on these experiments, spurious response at frequencies less than the half-sample rate (that is, in-band aliasing) has little effect on recognition or ID performance. It appears that some effect occurs at long ranges where acquisition probabilities are low; this is logical because, at long range, there are very few pixels on target. However, at ranges of practical interest, in-band corruption tends to affect minor details but does not change the basic presence or location of important cues.

Out-of-band spurious response, however, tends to mask the underlying image. Line raster, pixel edges, and other spurious high-frequency content does degrade targeting performance. The amount of performance degradation depends on the ratio of spurious content to image content. The spurious response ratio (SRR_{out}) of integrated out-of-band spurious response to the integrated transfer response is a good indicator of performance degradation.

$$SRR_{out} = \frac{\int_{\nu/2}^{\infty} \text{Spurious response}\, d\xi}{\int_{0}^{\infty} \text{Transfer response}\, d\xi} \tag{7.2}$$

Many imagers have different sample spacings horizontally and vertically; for example, most scanning thermal imagers have better sampling in the horizontal direction. SRR_{out} is calculated in the two dimensions independently, and the geometric mean is used to estimate performance degradation.

In real imagers, the display and eye MTF limit the frequency content visible to the observer. When doing numerical integrals, a practical limit for the upper frequency is 2.5

times the sample frequency. Also, the replicas centered on frequencies above twice the sample frequency are effectively filtered out. Quite often, the replicas of the pre-sample MTF overlap in the frequency domain; in Figure 7.1, there is a small overlap between the first and second replicas. In the overlap region, the signals from different replicas are root-sum-squared before integration.

$$SRRH_{out} = \frac{\int_{\nu/2}^{2.5\nu} \sum_{n=-2,-1,1,2} H_{pre}(\xi - n\nu) H_{post}(\xi) d\xi}{\int_{0}^{2.5\nu} H_{sys}(\xi) d\xi} \tag{7.3}$$

$$SRRV_{out} = \frac{\int_{\mu/2}^{2.5\mu} \sum_{n=-2,-1,1,2} V_{pre}(\eta - n\mu) V_{post}(\eta) d\eta}{\int_{0}^{2.5\mu} V_{sys}(\eta) d\eta} \tag{7.4}$$

When predicting the probability of accomplishing a task at range, sampling artifacts reduces the resolved cycles.

$$N_{sampled} = N_{resolved} \sqrt{1 - 0.58 SRRH_{out}} \sqrt{1 - 0.58 SRRV_{out}} \tag{7.5}$$

$N_{resolved}$ is the resolved cycles on target calculated using Equation 6.5. $N_{sampled}$ is used in lieu of $N_{resolved}$ in Equation 6.6 to calculate probability.

In these equations,

$SRRH_{out}$ = out-of-band spurious response ratio in horizontal dimension
$SRRV_{out}$ = out-of-band spurious response ratio in vertical dimension
$H_{pre}(\xi)$ = horizontal pre-sample MTF
$V_{pre}(\eta)$ = vertical pre-sample MTF
ξ = horizontal spatial frequency in (milliradian)$^{-1}$
η = vertical spatial frequency in (milliradian)$^{-1}$
ν = horizontal sample frequency in (milliradian)$^{-1}$
μ = vertical sample frequency in (milliradian)$^{-1}$
$H_{eye}(\xi \text{ or } \eta)$ = eyeball MTF
$H_{elec}(\xi)$ = horizontal electronics MTF
$V_{elec}(\eta)$ = vertical electronics MTF
$H_{dsp}(\xi)$ = horizontal display MTF
$V_{dsp}(\eta)$ = vertical display MTF
$H_{sys}(\xi)$ = horizontal system MTF
$V_{sys}(\eta)$ = vertical system MTF
$H_{post}(\xi) = H_{elec}(\xi)H_{dsp}(\xi)H_{eye}(\xi)$
$V_{post}(\eta) = V_{elec}(\eta)V_{dsp}(\eta)H_{eye}(\eta)$

7.2.1 Discussion

A sampling model which ignores corruption of the baseband signal would seem to be counter-intuitive. There *must be* a point at which the original signal is so corrupted by aliasing that a performance impact results. Experiments 36 and 44 were run to examine this case. These experiments are described in detail in Appendices A and C. Experiment 36 used an ID task and Experiment 44 used a recognition task. A large amount of aliasing at frequencies less than the half-sample frequency was created by using a very small detector fill factor. These experiments support the conclusion that range degradation is predicted based on the out-of-band spurious response.

As described in the appendices, F/2, diffraction limited optics were used with a 256 by 256 staring array which had a 0.0016 fill factor (one micron square detector on a 25 micron square pitch). The sampled imagery appeared corrupted; the internal details and shape of the target vehicles were significantly distorted. Intuitively, viewing the images, it appeared that scene structure was destroyed, not that raster or display pixel structure was obscuring the underlying scene details.

Nonetheless, experimental results support the conclusion that performance degradation due to sampling is predicted by the amount of out-of-band spurious response. This result might be more understandable when it is realized that the small detectors were generating large amounts of out-of-band energy; the in-band signal was being aliased in a way that created significant high frequency content that was not filtered out by even good display pixel interpolation. The small fill factor did result in a 27% loss in range performance, but the performance loss was predictable from the out-of–band Spurious Response Ratio.

All of the sampling experiments have involved either identifying or recognizing targets; the applicability of the model for the detection task has not been verified. As discussed in Appendices A and C, the sampling adjustment appears to be optimistic when the targets are at long ranges and have few samples on target. That is, sampling appears to have a greater effect on ID and recognition when the targets are at long range and are poorly resolved. This might infer that the detection task, which involves few cycles on target, is more affected by sampling than either ID or recognition.

It should be remembered, however, that recognition is the level of discrimination at which the observer knows he is looking at a target. The low cycle criteria associated with the detection task occurs because of the acceptance of many false alarms. It is possible that Equation 7.5 needs to be adjusted to accurately predict detection, but that is not certain. Based on experience using 1st generation thermal imagers in search experiments, poor vertical sampling lead to increased false alarms, not an increase in the number of cycles needed to detect the target.

8

Modeling Reflected-Light Imagers

Imagers of reflected light operate in the spectral band between 0.4 and 2.0 microns. This spectral region constitutes the visible band from 0.4 to 0.75 microns and the near infrared (NIR) band from 0.75 to 3.0 microns. Quite often, light with wavelengths between one and two microns is called short wave infrared (SWIR). Natural light is abundant in the 0.4 to 2.0 micron spectral band. Figure 8.1 shows illumination from sunlight, moonlight, and starlight (including airglow). The visible band is especially bright in the day, and the SWIR is the brightest of the three bands on a moonless night. The figure shows illumination through the atmosphere; the moon and sun are both at a 60 degree zenith angle. There are four distinct atmospheric absorption bands apparent in the illumination spectra; these are at 0.95, 1.1, 1.4, and 1.9 microns. These absorption bands also affect atmospheric transmission; transmission over a one kilometer, horizontal path is shown in Figure 8.2. In addition to abundant natural illumination, the clear atmosphere is fairly transparent over most of the 0.4 to 2.0 micron spectral band.

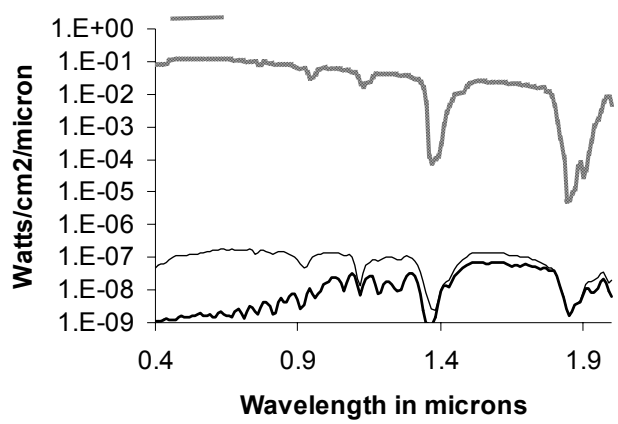


Figure 8.1 Illumination from the sun, moon, and starlight. Most of the "starlight" illumination is actually from air glow.

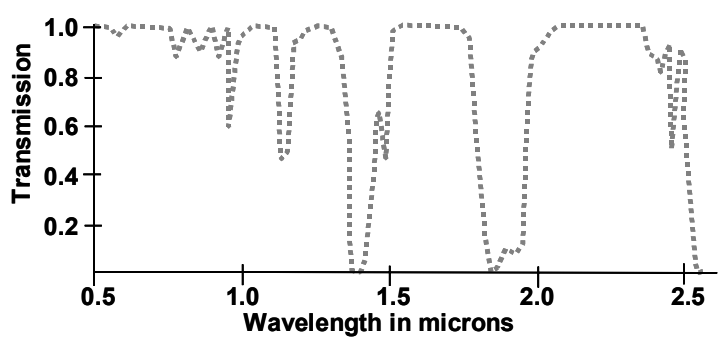


Figure 8.2 Atmospheric transmission over a one kilometer, horizontal path.

Target and background reflectivities tend to vary with wavelength in this spectral region; natural and manmade objects tend not to be gray bodies. Figure 8.3 shows the spectral reflectivity of a foreign paint, sand, gravel, a mixed soil, and dead grass. The paint closely matches the gravel and soil out to about 1.2 microns and closely matches the sand beyond 1.2 microns. The paint has very different reflectivity properties from dead grass (the top curve in the figure) over the entire spectral range. The apparent contrast seen by the imager depends on the background and also on the spectral band chosen

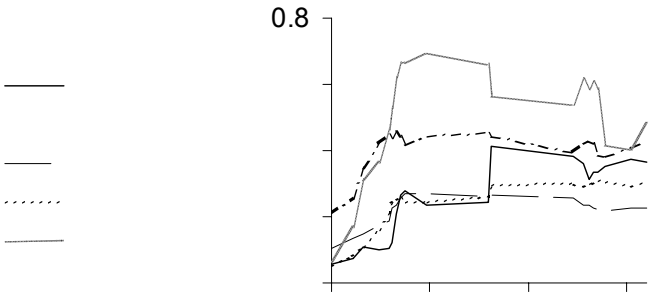


Figure 8.3 Spectral reflectivities of a foreign, tactical-vehicle paint and various kinds of dirt and grass.

8.1 Staring Focal Plane Arrays

The theory for a solid state camera is developed in this section. A diagram of a solid state imager is shown in Figure 8.4. A lens focuses light onto a two-dimensional focal plane array of detectors (the FPA). Photo-current is generated in each detector for a fraction of each frame or field interval; the stored charge is read out and formatted for display.

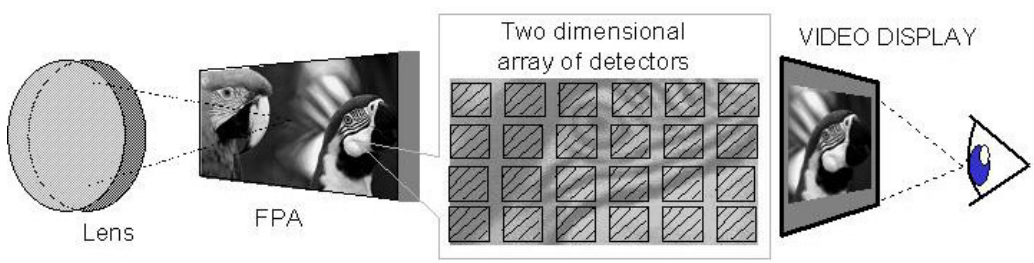


Figure 8.4 Diagram of a solid state imager.

The calculation of photo current is described in references such as the *Electro-Optics Handbook* (Burle Industries, 1974). The detector current from a scene element is calculated as follows.

$$photocurrent = \frac{V_{det}\, H_{det}}{4F\#^2} \int_0^\infty I(\lambda)\, T(\lambda)\, R(\lambda)\, R_{sp}(\lambda)\, C(\lambda)\, d\lambda \qquad (8.1)$$

where

λ = Wavelength in μm

F_0 = focal length of objective lens in centimeters

F# = focal length F_O divided by aperture diameter

$I(\lambda)$ = Illumination in watts cm^{-2} $micron^{-1}$

$T(\lambda)$ = Transmission of atmosphere as a function of λ

$R(\lambda)$ = Spectral reflectance of the scene element as a function of λ

$R_{sp}(\lambda)$ = Detector response in amperes per watt as a function of λ

$C(\lambda)$ = Objective lens and spectral filter transmission as a function of λ

V_{det} and H_{det} are vertical and horizontal dimensions of detector active area in centimeters

Let R_T and R_B represent the detector photocurrent spectral integral for targets and backgrounds, respectively. Because the signal is proportional to the photo-current and noise is proportional to the square root of the photo-current, the average electron flux per solid angle is used in the model. The spatial frequency unit is cycles per milliradian. We want to calculate the average number of electrons per second in a square milliradian (E_{av}); this is because noise power spectral density has units of $(\text{second-milliradian}^2)^{-1}$. Power spectral density is in the frequency domain; the calculation here is in the space domain.

$$E_{av} = 0.5\ 10^{-6}\ (R_T+R_B)\ F_O^2 / (V_{pit}\ H_{pit}\ e^-) \tag{8.2}$$

where

e^- = charge on an electron (1.6E-19 Coulombs per electron)

H_{pit} and V_{pit} are horizontal and vertical detector pitch in centimeters

The ratio $F_O^2 / (V_{pit}H_{pit})$ gives the number of photo-detectors in a square radian; the E-6 factor converts this to the number in a square milliradian. The unit "square radian" rather than steradian seems strange; remember, however, that the model treats two dimensions as a one-dimensional calculation, done twice. Calculations are not really done in two-dimensional space.

Equations 4.14 and 4.15 for $CTFH_{sys}$ and $CTFV_{sys}$ can now be written for a solid state imager. In the following equations, ξ' and η' are dummy variables of integration.

$$CTFH_{sys}(\xi) = \frac{CTF\left(\xi / SMAG\right)}{M_{dsp} H_{sys}(\xi)} \left(\frac{1}{\kappa_{con}^2} + \frac{\alpha^2 E_{av}\, QH_{hor}(\xi) QV_{hor}}{E_{av}^2} \right)^{1/2} \tag{8.3}$$

$$CTFV_{sys}(\eta) = \frac{CTF\left(\eta / SMAG\right)}{M_{dsp} V_{sys}(\xi)} \left(\frac{1}{\kappa_{con}^2} + \frac{\alpha^2 E_{av}\, QH_{ver} QV_{ver}(\eta)}{E_{av}^2} \right)^{1/2} . \tag{8.4}$$

where

α = 169.6 (a proportionality constant)

ξ = horizontal spatial frequency in $(\text{milliradian})^{-1}$

η = vertical spatial frequency in $(\text{milliradian})^{-1}$

CTF(ξ/SMAG) = naked eye Contrast Threshold Function; see Appendix E
κ_{con} = contrast enhancement
B(ξ or η) = the Equation (3.9) eye filters
H_{eye}(ξ or η) = eyeball MTF; see Appendix E
H_{elec}(ξ) = horizontal electronics MTF
V_{elec}(η) = vertical electronics MTF
H_{dsp}(ξ) = horizontal display MTF
V_{dsp}(η) = vertical display MTF
H_{sys}(ξ) = horizontal system MTF
V_{sys}(η) = vertical system MTF
QH_{hor} = horizontal noise bandwidth for $CTFH_{sys}$ defined by Equation 8.5
QV_{hor} = vertical noise bandwidth for $CTFH_{sys}$ defined by Equation 8.6
QH_{ver} = horizontal noise bandwidth for $CTFV_{sys}$ defined by Equation 8.7
QV_{ver} = vertical noise bandwidth for $CTFV_{sys}$ defined by Equation 8.8

$$QH_{hor}(\xi) = \int \left| B(\xi'/\xi) H_{elec}(\xi') H_{dsp}(\xi') H_{eye}\left(\xi'/SMAG \right) \right|^2 d\xi' \tag{8.5}$$

$$QV_{hor} = \int \left| V_{elec}(\eta) V_{dsp}(\eta) H_{eye}\left(\eta/SMAG \right) \right|^2 d\eta \tag{8.6}$$

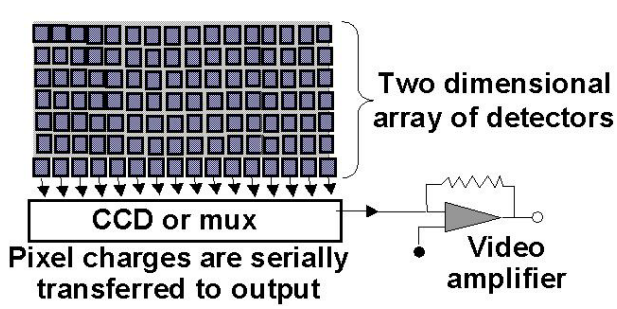
$$QH_{ver} = \int \left| H_{elec}(\xi) H_{dsp}(\xi) H_{eye}\left(\xi/SMAG \right) \right|^2 d\xi \tag{8.7}$$

$$QV_{ver}(\eta) = \int \left| B(\eta'/\eta) V_{elec}(\eta') V_{dsp}(\eta') H_{eye}\left(\eta'/SMAG \right) \right|^2 d\eta' \tag{8.8}$$

Equations 8.3 and 8.4 assume ideal shot noise; other noise sources are ignored. This assumption is realistic for most cameras under high illumination conditions. However, as the light fails, noise sources other than shot noise begin to dominate.

Figure 8.5 illustrates the read-out of a CCD imager. Photo-charges are clocked down, line by line, until they reach the horizontal shift register. After each line is entered into the register, it is shifted out at high speed through the video amplifier. In this manner, the imagery collected in parallel at each detector becomes a serial stream. The benefit is a single output line, generally formatted as RS-170 standard video. The penalty is that the high speed video amplifier is noisy.

Figure 8.5 Diagram of Video Read-out High bandwidth video amplifier adds noise to the signal.



The video amplifier noise is typically specified in terms of noise electrons per pixel per field or frame. Although the noise actually arises in the amplifier or read-out circuitry, manufacturers provide the equivalent number of noise electrons in order to make calculation of dynamic range and total noise easier.

A second common source of excess noise is dark current. Dark current is often specified as electrons per pixel per frame. Sometimes, dark current is specified as current density; for example, the dark current might be specified as 100 microamperes per square centimeter. In that case, the active detector area and frame time are used to calculate dark electrons per pixel per frame. The noise associated with dark current is the square root of the number of dark current electrons.

All noise sources are added in quadrature. The noise in one second and one square milliradian is:

$$E_{noise} = E_{av} + \frac{T_{CCD}\left(E_{amp}^2 + E_{DC}\right)F_0^2}{H_{pit}V_{pit}}10^{-6} \tag{8.9}$$

where
E_{amp} = the amplifier noise in electrons per pixel per frame
E_{DC} = dark current electrons per pixel per frame
T_{CCD} = fields or frames per second.

The equations for threshold vision through the imager now become:

$$CTFH_{sys}(\xi) = \frac{CTF\left(\xi/SMAG\right)}{M_{dsp}H_{sys}(\xi)}\left(\frac{1}{\kappa_{con}^2} + \frac{\alpha^2 E_{noise}\,QH_{hor}(\xi)QV_{hor}}{E_{av}^2}\right)^{1/2} \tag{8.10}$$

$$CTFV_{sys}(\eta) = \frac{CTF\left(\eta/SMAG\right)}{M_{dsp}V_{sys}(\xi)}\left(\frac{1}{\kappa_{con}^2} + \frac{\alpha^2 E_{noise}\,QH_{ver}QV_{ver}(\eta)}{E_{av}^2}\right)^{1/2}. \tag{8.11}$$

Since amplifier noise can completely dominate performance at low illumination levels, techniques have been developed to provide signal gain prior to the read-out electronics. Generally, however, the electron gain is non-ideal in the sense that the gain itself generates excess noise. Sometimes the amount of excess noise depends on the gain applied. For example, avalanche silicon diodes have excess noise equal to the square root of the gain; a gain of 100 comes at the cost of increasing shot noise by a factor of 10.

Let N_f represent the noise factor which is always greater than one. N_f might be a fixed value or might depend on gain through the equation

$$N_f = Gain^{\gamma}$$

where γ is an exponent which depends on the technology used; for silicon avalanche diodes, γ is 0.5. Then E_{noise} in Equations 8.10 and 8.11 becomes:

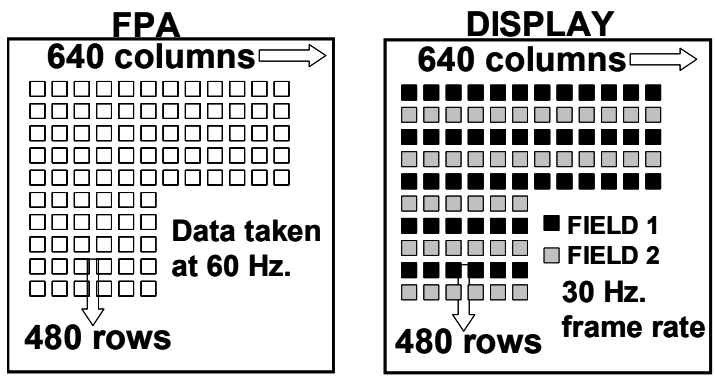
$$E_{noise} = \left(E_{av} + \frac{T_{CCD} E_{DC} F_0^2}{H_{pit} V_{pit}} 10^{-6} \right) N_f^2 + \frac{T_{CCD} E_{amp}^2 F_0^2}{Gain^2 H_{pit} V_{pit}} 10^{-6} \qquad (8.12)$$

If both gain and noise factor are unity, then Equation 8.12 reduces to 8.9.

8.1.1 Interlace

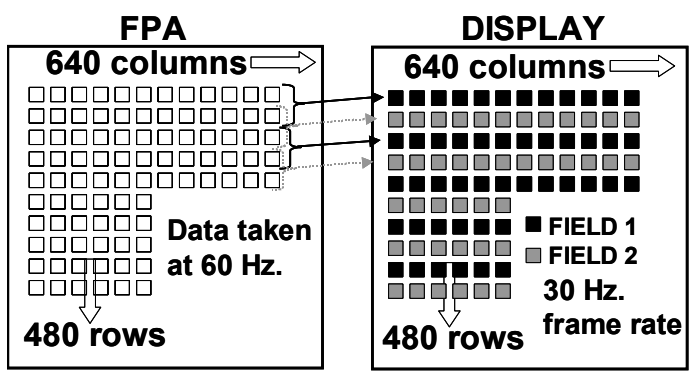
Display interlace is used to reduce electronic bandwidth while maintaining a high resolution image. Electronic interlace, also called standard interlace or simply interlace, is illustrated in Figure 8.6. The FPA operates at 60 Hertz. However, the display operates at a 30 Hertz frame rate. The first, third, fifth, every odd line from the FPA is displayed in the first field. The even lines (two, four, six, etcetera) are displayed in the second field. Although interlace does not degrade resolution, the displayed signal to noise is affected because half the available signal from the FPA is discarded.

Figure 8.6 Illustration of Electronic Interlace



Pseudo interlace is a means for using all of the signal electrons while maintaining the reduced bandwidth benefits of interlace. In the first display field, photo-electrons from pixels in rows one and two are added and presented on display line 1. Pixels on lines three and four are added and presented on display line 3. The process continues, adding odd lines to even lines and displaying on odd lines. In field two, FPA lines two and three are added and presented on display line 2. Even FPA lines are added to odd lines and displayed on the even lines. This process is illustrated in Figure 8.7. Pseudo interlace uses all of the available signal electrons and therefore maintains image sensitivity. Also, field alignment is properly maintained; samples are in the correct location. The penalty paid is a decrease in the vertical MTF of the imager.

Figure 8.7 Illustration of Pseudo Interlace



In Equations 8.10 and 8.11, E_{av} is divided by two for standard interlace but is not affected by pseudo interlace. E_{noise} in Equation 8.12 is affected as shown in Equation 8.13 where I_{signal} and I_{amp} are defined:

$I_{amp} = I_{signal} = 1$ for non-interlace
$I_{amp} = 2$ for any interlace
$I_{signal} = 2$ for electronic interlace and
$= 1$ for pseudo interlace.

$$E_{noise} = \left(E_{av} + \frac{T_{CCD} E_{DC} F_0^2}{H_{pit} V_{pit}} 10^{-6} \right) \frac{N_f^2}{I_{signal}} + \frac{T_{CCD} E_{amp}^2 F_0^2}{Gain^2 H_{pit} V_{pit}} \frac{10^{-6}}{I_{amp}} \tag{8.13}$$

8.1.2 Snapshot and Frame Integration

Temporal integration of the eye varies with light level. As illumination decreases, the eye integrates for a longer period. If detector noise is temporally varying at a fairly rapid rate (50 or 60 imager fields per second is adequate), then the eye temporally filters detector noise in the same way as eye noise. However, if a snapshot (single frame) is taken, or if frame integration is used, then the effect of eye integration time must be explicitly considered.

The dependence of eye integration time on display luminance is:

$t_{eye} = .0192 + .0625\ (L / 1.076)^{-.17}$

where L is display luminance and t_{eye} is integration time.

For snapshot imagery, define t_{act} as:

t_{act} = frame time for non-interlace and pseudo interlace

= half of a frame time for electronic interlace.

Then E_{noise} for snapshot ($E_{noi\text{-}snap}$) is related to E_{noise} for framing shown in Equation 8.9 by:

$$E_{noi-snap} = \frac{t_{eye}}{t_{act}} E_{noise} \tag{8.14}$$

If frame integrate is used, then the effect depends on whether the imager is in framing or snapshot mode. If in snapshot mode, then the benefit of integrating F_{INT} frames is:

$$E_{noi-snap} = \frac{t_{eye}}{F_{INT}\, t_{act}} E_{noise} \tag{8.15}$$

If the imager is in framing mode, then the benefit of frame integration is moderated by the fact that the eye is already integrating temporally.

$$E_{frm-\text{int}} = \frac{t_{CCD}\, t_{eye}}{(F_{INT} + t_{eye}\, t_{CCD})} E_{noise} \tag{8.16}$$

8.2 Direct View Image Intensifiers

Image intensifiers amplify moonlight and starlight at spectral wavelengths between 0.5 and 1.0 micron. To the left in Figure 8.8, a pilot is wearing the Aviator's Night Vision Imaging System (ANVIS) which consists of two oculars, one for each eye. A schematic of one direct view goggle ocular is shown at right. The objective lens forms an inverted image of the scene on the image intensifier tube. The I^2 tube amplifies the brightness of the image as described below. The fiber-optic twist erects the brighter image. The eyepiece creates a unity magnification, virtual image of the scene, allowing the pilot to fly at night without lights. By modifying the eyepiece to create image magnification, a single ocular can also be an effective rifle sight.

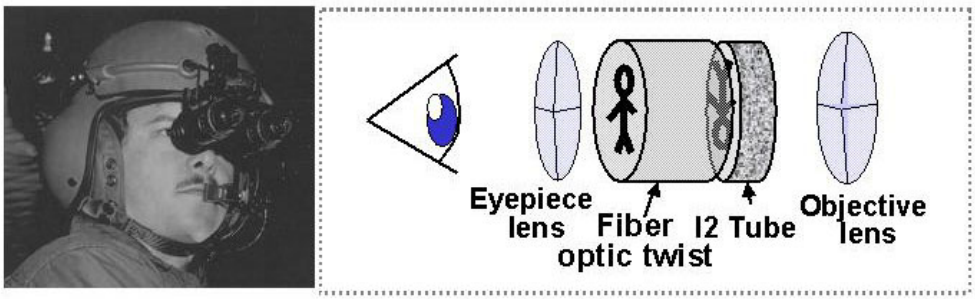
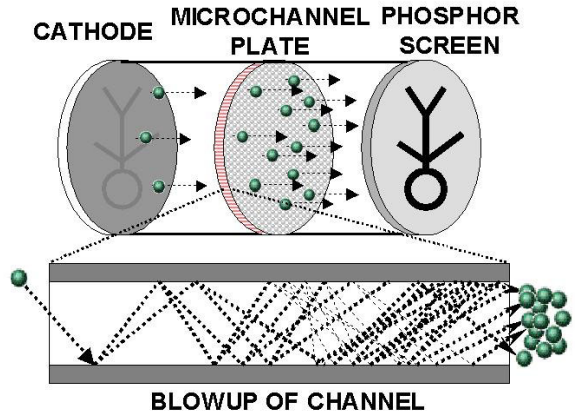


Figure 8.8 ANVIS goggle shown at left; at right is a schematic diagram of a single ocular.

Operation of the I^2 tube is illustrated in Figure 8.9. Photons from the scene generate photo-electrons in the cathode. A high voltage accelerates the photo-electrons to the micro channel plate (MCP). The MCP consists of millions of tiny channels; these channels are about five microns in diameter on a pitch of six microns. The channel length to diameter ratio is about seventy. Operation of the MCP is shown by the blowup of a single channel at the bottom of the figure. Photo-electrons enter the channel and are accelerated by a high voltage across the channel plate. Secondary electrons are emitted when the photo-electrons strike the channel wall. The secondary electrons are then accelerated, strike the wall, and create more electrons. Electron gain through the channel is controlled by varying the voltage across the MCP. Channel electrons exit the MCP and are accelerated by another high voltage to the phosphor where an image is formed. Brightness gain results from the MCP electron gain, the energy gained from electron acceleration between the MCP and phosphor, and from the fact that the cathode is sensitive to a much broader range of light wavelengths than the eye.

Brightness gain is specified by the ratio foot Lamberts from the phosphor to foot candles on the cathode. Typical gain is 30,000 but gains to 100,000 are possible; however, excessive gain leads to bothersome scintillation in the image. Brightness output of the I^2 tube is controlled by limiting the current available to the MCP; generally, goggle brightness is limited to about 3 fL. Tube noise factor is ideally 1.4 based on the open area ratio (not all photo-electrons get through the MCP); typical noise factor is about two. See Bender (2000) for a more thorough discussion of the theory and specification of image intensifiers.

Figure 8.9 Illustration of I^2 Tube Operation



Equation 8.17 is used to find the photocurrent for one square centimeter of cathode area.

$$photocurrent = \frac{1}{4F\#^2}\int_{-\infty}^{\infty} I(\lambda)T(\lambda)R(\lambda)R_{sp}(\lambda)C(\lambda)\,d\lambda \tag{8.17}$$

$$E_{av} = 3.13E12\ (R_T+R_B)\ F_O^2\ B_{io} \tag{8.18}$$

In Equation 8.18, R_T and R_B are the photocurrent integrals in amperes per square centimeter for target and background, respectively. The 3.13E12 factor is the product of 0.5 to average target and background flux, 1E-6 to convert radians to milliradians, and divide by the charge on an electron (1.6E-19 Coulombs per electron). The B_{io} factor accounts for the improved signal to noise available from systems with two image intensifier tubes.

B_{io} = 1 for monocular or biocular (one I^2 tube)
= 2 for binocular (two I^2 tubes)

In image intensifiers, dark current is called equivalent background input (EBI). Although generally not important at room temperature, EBI can be significant in very hot environments or if the I^2 tube is enclosed with other hardware. The unit for EBI is foot candles (lumen per square foot) of 2856 K light. Tube specification sheets generally list the responsivity of the tube (R_{esp}) in uamps per lumen of 2856K light. So the dark current (DC_{EBI}) per square centimeter of cathode area is:

$$DC_{EBI} = EBI\ R_{esp}\ 1E\text{-}6 / 929.03 \tag{8.19}$$

where the 929.03 factor converts square feet to square centimeters. The noise electrons (E_{noise}) in one second and one square milliradian is:

$$E_{noise} = E_{av} + 6.25E6\ DC_{EBI}\ F_O^2\ B_{io} \tag{8.20}$$

In order to establish the eye CTF, the output brightness of the tube (B_{out}) must be calculated. First the current to light gain (G_{elec} fL cm^2/uamp) is calculated from a knowledge of tube gain (G_{tube}) in fL/fc and tube responsivity.

$$G_{elec} = G_{tube}\ 929.03 / R_{esp} \tag{8.21}$$

For an eyepiece transmission of $\tau_{eye,}$ the output brightness is:

$$Bout = 0.5(R_T + R_B)\ G_{elec}\ \tau_{eye} + EBI\ G_{tube}\ \tau_{eye} \tag{8.22}$$

The equation for horizontal and vertical threshold vision through the imager is:

$$CTF_{sys}(\xi) = \frac{CTF\left(\xi / SMAG\right)}{(E_{av} / E_{noise}) H_{sys}(\xi)} \left(\frac{1}{\kappa_{con}^2} + \frac{\alpha^2 N_f^2 E_{noise}\, QH(\xi) QV}{E_{av}^2} \right)^{1/2} \quad (8.23)$$

where E_{av} / E_{noise} is the contrast degradation M_{dsp} due to EBI. For direct view I2 systems, MTF loss is associated with the optics (H_{opt}), the tube (H_{tube}), and the eyepiece (H_{ep}). Since little of the tube MTF is associated with the cathode, tube MTF filters the noise.

$$H_{sys} = V_{sys} = H_{opt}\, H_{tube}\, H_{ep} \quad (8.24)$$

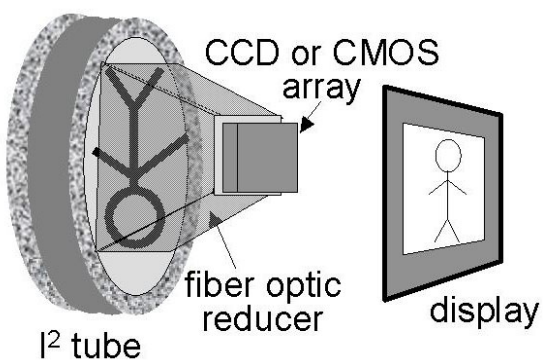
$$QH(\xi) = \int \left| B(\xi'/\xi) H_{tube}(\xi') H_{ep}(\xi') H_{eye}\left(\xi' / SMAG\right) \right|^2 d\xi' \quad (8.25)$$

$$QV = \int \left| H_{tube}(\eta) H_{ep}(\eta) H_{eye}\left(\eta / SMAG\right) \right|^2 d\eta \quad (8.26)$$

8.3 I^2 Optically Coupled to CCD or CMOS

The eyepiece in Figure 8.8 can be replaced by a CCD or CMOS imager and display; this allows the image intensifier to be mounted remotely from the observer. A fiber-optic minifier or optical relay lens is used because the image intensifier format is generally twice as large as the CCD image array. See Figure 8.10 for a schematic diagram of an I^2 CCD camera.

Figure 8.10 Illustration of I^2 Tube Optically Coupled to CCD or CMOS FPA



The MTF of the CCD and display is applied to the I^2 tube signal and noise. H_{CCD} and V_{CCD} are the horizontal and vertical CCD MTF, respectively. The CCD noise is filtered by the CCD MTF, display MTF, the eyeball, and the perceptual filter. The CCD noise is added in quadrature with the other noise terms; this means that CCD noise must be expressed in terms of cathode photoelectrons.

$$E_{CCD} = \frac{\kappa_{cst}\left(E_{amp}{}^2 + B_{out} / R_{CCD}\right) R_{CCD}{}^2\, T_{CCD}\, A_{pix}\, K_h\, K_v\, F_o{}^2\, R_{esp}{}^2}{\left(G_{tube}\ e^-\right)^2} \quad (8.27)$$

where

E_{CCD} = CCD noise expressed as I^2 cathode photo-electrons
κ_{cst} = 4.53E13 = constant factors (charge on electron, units convrsion)
E_{amp} = Amplifier noise per CCD pixel per field in electrons,
T_{CCD} = field rate of CCD
R_{CCD} = footcandle to generate one electron in a CCD pixel each second
A_{pix} = area of a CCD pixel = H_{pit} V_{pit}
K_h and K_v are horizontal and vertical reduction ratios
B_{out} = light output from I^2 tube in fL

The shorthand "CCD" is used to represent the array, but the technology used for the solid state imager is not relevant. Also, optical coupling can be by coherent, fiber optic reducer as shown in the figure, or a relay lens can be used.

All calculations for the I^2 tube remain the same as in Equations 8.18 through 8.26 except for the addition of CCD and display MTF. The noise now has two terms because CCD noise is filtered differently than I^2 tube noise. EH_{hor} and EV_{hor} are the spatial filters for calculating horizontal CCD noise. EH_{ver} and EV_{ver} are the spatial filters for calculating vertical CCD noise. In the following, ξ' and η' are dummy variables of integration.

$$H_{sys} = H_{opt}\, H_{tube}\, H_{ep}\, H_{CCD}\, H_{dsp} \tag{8.28}$$

$$V_{sys} = H_{opt}\, H_{tube}\, H_{ep}\, V_{CCD}\, V_{dsp} \tag{8.29}$$

$$QH_{hor}(\xi) = \int \left| B(\xi'/\xi) H_{tube}(\xi') H_{ep}(\xi') H_{CCD}(\xi') H_{dsp}(\xi') H_{eye}\left(\xi'/SMAG \right) \right|^2 d\xi' \tag{8.30}$$

$$QV_{hor} = \int \left| H_{tube}(\eta) H_{ep}(\eta) V_{CCD}(\eta) V_{dsp}(\eta) H_{eye}\left(\eta/SMAG \right) \right|^2 d\eta \tag{8.31}$$

$$EH_{hor}(\xi) = \int \left| B(\xi'/\xi) H_{CCD}(\xi') H_{dsp}(\xi') H_{eye}\left(\xi'/SMAG \right) \right|^2 d\xi' \tag{8.32}$$

$$EV_{hor} = \int \left| V_{CCD}(\eta) V_{dsp}(\eta) H_{eye}\left(\eta/SMAG \right) \right|^2 d\eta \tag{8.33}$$

$$CTFH_{sys}(\xi) = \frac{CTF\left(\xi/SMAG\right)}{M_{dsp} H_{sys}(\xi)} \left(\frac{1}{\kappa_{con}^2} + \frac{\alpha^2 N_f^2 E_{noise}\, QH_{hor}(\xi) QV_{hor}}{E_{av}^2} + \frac{\alpha^2 E_{CCD}\, EH_{hor}(\xi) EV_{hor}}{E_{av}^2} \right)^{1/2} \tag{8.34}$$

$$QH_{ver} = \int \left| H_{tube}(\xi) H_{ep}(\xi) H_{CCD}(\xi) H_{dsp}(\xi) H_{eye}\left(\xi/SMAG \right) \right|^2 d\xi \tag{8.35}$$

$$QV_{ver}(\eta) = \int \left| B(\eta'/\eta) H_{tube}(\eta') H_{ep}(\eta') V_{CCD}(\eta') V_{dsp}(\eta') H_{eye}\left(\eta'/SMAG \right) \right|^2 d\eta' \tag{8.36}$$

$$EH_{ver} = \int \left| H_{CCD}(\xi) H_{dsp}(\xi) H_{eye}\left(\xi / SMAG \right) \right|^2 d\xi \tag{8.37}$$

$$EV_{ver}(\eta) = \int \left| B(\eta'/\eta) V_{CCD}(\eta') V_{dsp}(\eta') H_{eye}\left(\eta' / SMAG \right) \right|^2 d\eta' \tag{8.38}$$

$$CTFV_{sys}(\eta) = \frac{CTF\left(\eta / SMAG \right)}{M_{dsp} V_{sys}(\eta)} \left(\frac{1}{\kappa_{con}^2} + \frac{\alpha^2 N_f^2 E_{noise}\, QH_{ver}\, QV_{ver}(\eta)}{E_{av}^2} + \frac{\alpha^2 E_{CCD}\, EH_{ver}\, EV_{ver}(\eta)}{E_{av}^2} \right)^{1/2} \tag{8.39}$$

Vollmerhausen (1996) provides three validation examples; these show a good match between model predictions and experimental data. Appendix E provides details on how to model CCD MTF and fiber-optic taper MTF, respectively.

It is important to realistically assess display performance when modeling I^2 CCD cameras. This is particularly true when modeling low illumination levels, because the camera electronic output might not be sufficient to properly drive the display. As a result, the best "operator optimized" image might have poor display contrast. During the validation experiments described in Vollmerhausen (1996), operator-selected display contrast ranged between 10 percent and 40 percent when the various cameras were used under overcast starlight illumination.

Under low target-illumination conditions, some I^2 CCD cameras will output only millivolts of video signal. Cathode ray tubes have a power law relationship between the input voltage and the output luminance. At maximum gain and with no brightness control offset, typical displays will provide very little output luminance when the input voltage is only a few millivolts. Typical gamma correction circuits do not correct inputs this low.

Adding display brightness with the brightness control will move the image up the power law curve, providing a larger luminance change for a given change in input voltage. Adding brightness will also make the whole display brighter, improving the human visual response. As a result of the two properties together, the display might have the best subjective appearance with minimum display luminance greater than zero. The operator will choose poor contrast over no or very low luminance. Since, in this instance, CTF is inversely proportional to display contrast, the display characteristics can be a dominant factor in determining system performance.

8.4 CCD or CMOS Array inside I^2 Tube

The array can be inside the vacuum of the image intensifier tube. Electrons are directly gathered by the CCD rather than optically coupling the CCD to the I^2 tube phosphor output. This is illustrated in Figures 8.11 and 8.12. In 8.11, electrons are accelerated from the cathode to the CCD by a high voltage. The photo-electrons are given sufficient energy to create 100 to 200 secondary electrons when the CCD silicon is struck. This provides near-ideal electron gain. In Figure 8.12, an MCP is used. The MCP adds complexity but provides advantages. The MCP provides gain control; the cathode to CCD voltage in Figure 8.11 cannot be lowered too much, or the image will blur. Also,

with the MCP, secondary electrons at the CCD are not necessary; the CCD (or CMOS array) is just a collector of electrons and need not provide gain. The arrangement in Figure 8.12 is expected to prolong CCD array lifetime.

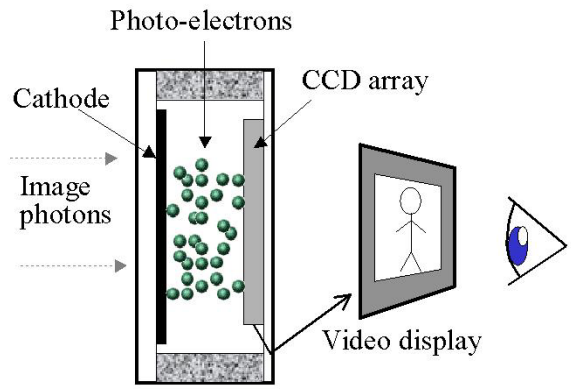


Figure 8.11 CCD array inside I^2 tube vacuum.

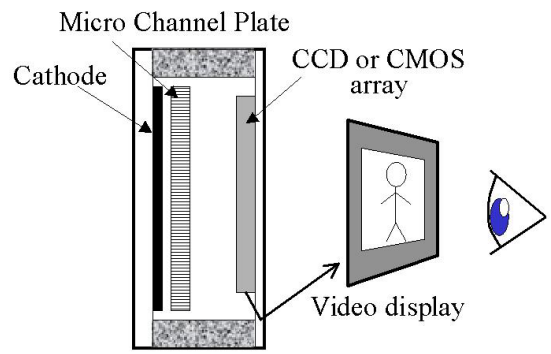


Figure 8.12 CCD array inside I tube vacuum; this arrangement also has an MCP.

CCD noise E_{CCD} is calculated differently from that used with optically coupled arrangement. Otherwise, Equations 8.30 through 8.39 are used to calculate CTF for these imagers. Using the same R_T and R_B as in Equation 8.18 for the photo-current per square centimeter of cathode area, the E_{CCD} is:

$$E_{CCD} = \frac{1E-6\left(E_{amp}{}^2 + 0.5(R_T + R_B)H_{\det}V_{\det}G_{elec}\right)T_{CCD}\,F_o{}^2}{H_{pit}V_{pit}\,G_{elec}^2} \tag{8.40}$$

where G_{elec} is the electron gain. This E_{CCD} is used in Equations 8.34 and 8.39.

8.5 Predicting Probability versus Range

8.5.1 Contrast Transmission through the Atmosphere

When predicting contrast transmission, certain assumptions are made to simplify calculations. These assumptions constrain the scenario for which the model is appropriate.

a) The target and background are co-located; the target is viewed against local terrain. Range to the target and range to the background are the same. From a

military standpoint, this is not an unreasonable assumption, and it relieves the necessity to consider some complex situations where target to background contrast can actually reverse.

b) Contrast loss through the atmosphere is from scattering. Contrast is not affected by absorption in the atmosphere. As shown in Figures 8.2 and 8.3, the atmospheric absorption bands remove light from the illumination. Most of the atmospheric path occurs before the light hits the target and background. Atmospheric absorption is considered when predicting spectral illumination.

c) Average luminance seen by the imager does not change with range. Target-to-background signal disappears into the average luminance established by the target and background reflectivities. CTF_{sys} (or MRC) depends on the light entering the sensor; noise, for example, is proportional to the square root of average luminance. In order to use a single, pre-calculated CTF_{sys} to represent imager performance, the assumption must be made that luminance does not change with range.

d) Contrast is reduced by scattering of target signal out of the line of sight and by sunlight, moonlight, or starlight scattered by the atmosphere into the imagers field of view. See Figure 8.13 for an illustration. In most scenarios, path radiance caused by light scattered into the sensor's path is the most serious cause of target–to-background contrast loss. The atmospheric path can appear brighter at the imager than the zero range target and background; this results in substantial loss of contrast. This part of the model is not completely self-consistent, since the luminance viewed by the imager is increasing with range under these circumstances. However, the approximation that the luminance is constant does not generally lead to serious errors. The most important factor is that contrast is greatly reduced by the atmosphere.

e) Path radiance is quantified by the Sky-to-Ground-Ratio (SGR). As the atmospheric path lengthens, the path becomes brighter. At some point, the path becomes "optically thick." That is, only light from the path is seen, and increasing the path length does not change the path radiance because as much light is scattered out as in. The SGR is the ratio between the maximum path radiance and the zero range radiance. SGR does not vary with range because the peak, long range value is used in the ratio. Table 8.1 gives values of SGR for a range of environments. Figure 8.14 shows the effect of SGR on contrast transmission. Equation 8.41 is used to calculate contrast loss for range R_{ng}, Beer's Law coefficient β, and target zero-range contrast C_{TGT-0}.

$$C_{TGT} = \frac{C_{TGT-0}}{1 + SGR\left(\exp^{-\beta R_{ng}} - 1\right)} \tag{8.41}$$

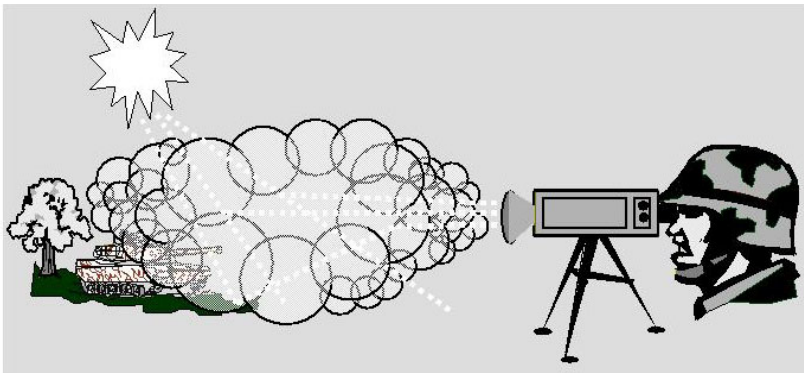


Figure 8.13 Sunlight scattered from atmosphere degrades target-to-background contrast.

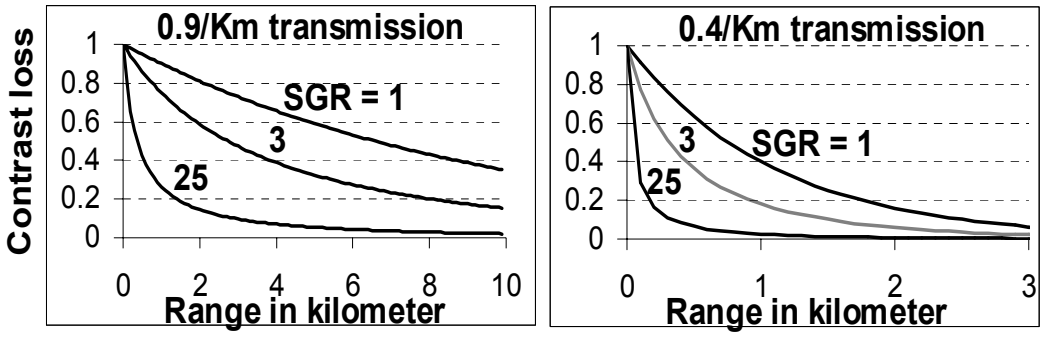


Figure 8.14 Effect of SGR on Contrast Transmission Left shows effect when the Beer's Law transmission is 0.9 per Kilometer. Right shows effect with 0.4 per Kilometer transmission.

Table 8.1 Typical Values for SGR

Typical SGR Values		
Terrain	*Clear*	*Overcast*
Desert	**1.4**	**7.0**
Forest	**5**	**25**
Typical	**3.0**	

8.5.2 Effect of Contrast Enhancement

Looking at examples like Equations 8.3, 8.4, 8.34, and 8.39, each $CTFH_{sys}$ and $CTFV_{sys}$ has two or more terms. One of the terms represents eye contrast limitations and depends on κ_{con}; the other term(s) depend on sensor noise and are independent of κ_{con}. In use, an imager may or may not have the contrast optimized to view the target, so contrast enhancement is one option that can be changed when calculating probability versus range. During a search, for example, the sensor is set to see the environment; the target has not been found. When a likely location is found, however, then the imager might be optimized to see if a target is present. So contrast enhancement might not be used in the wide field of view during search, but would be employed for the target identification task.

The process is the same for $CTFH_{sys}$ and $CTFV_{sys}$, so horizontal calculations are used as examples. Using Equation 8.3 for $CTFH_{sys}$:

$$CTFH_{sys}(\xi) = [CTFH_{eye}^2(\xi) + CTFH_{sen}^2(\xi)]^{1/2} \tag{8.42}$$

where

$$CTFH_{eye}(\xi) = \frac{CTF\left(\xi / SMAG\right)}{\kappa_{con} M_{dsp} H_{sys}(\xi)} \tag{8.43}$$

$$CTFH_{sen}(\xi) = \frac{CTF\left(\xi / SMAG\right)}{M_{dsp} H_{sys}(\xi)} \left(\frac{\alpha^2 E_{av} QH_{hor}(\xi) QV_{hor}}{E_{av}^2} \right)^{1/2} \tag{8.44}$$

Similar equations can be written for $CTFV_{sys}$. Models like SSCAM, I2MRC, and I2CCD output four arrays for each illumination and target-background combination modeled. Those arrays are $CTFH_{eye}$, $CTFH_{sen}$, $CTFV_{eye}$, and $CTFV_{sen}$. The value for κ_{con} determines how the four arrays are combined to predict the system CTF.

The options provided in the reflective model are: no contrast enhancement ($C_{dsp} = C_{TGT}$), display contrast of 0.25 ($C_{dsp} = 0.25$), and display contrast of 0.5 ($C_{dsp} = 0.5$). The value of 0.5 was determined by optimizing a set of 144 tactical vehicle image (twelve aspects of twelve different vehicles). Each image was individually optimized to bring out the particular cues needed to ID that vehicle at that aspect. Linearity was not enforced; pieces of the picture were subdued or enhanced as necessary to provide an optimum for identification. Once the optimizing process was complete, the contrast of the set was measured at 0.5. We feel that it is impossible for an automated process to duplicate this degree of optimization, and that 0.5 therefore represents an extreme for modeling purposes.

The 0.25 option resulted from applying histogram equalization, local area processing, and allowing some non-linear suppression of bright areas. The process was "by hand" in the sense that we ensured that no cues were lost due to the histogram equalization placement of gray levels. The measured contrast of the resulting target set was 0.25. This represents the contrast that can probably be achieved automatically.

When doing range calculations,

$$\kappa_{con} = \frac{C_{dsp}}{C_{TGT}} \tag{8.45}$$

As range to the target increases and target contrast (C_{TGT}) decreases, contrast enhancement maintains the displayed contrast at a high level. Of course, while the eye term in CTF_{sys} can be moderated by contrast enhancement, the noise term cannot. Noise must be low for contrast enhancement to help range performance.

8.5.3 Calculating Probability of Task Performance

At each range, apparent contrast C_{TGT} is established based on zero range contrast $C_{TGT\text{-}0}$ by using Beer's Law or MODTRAN. Contrast enhancement model is selected, and then κ_{con} is calculated using Equation 8.45. $CTFH_{sys}$ and $CTFV_{sys}$ can then be calculated. The TTP metric is calculated for both the horizontal and vertical dimensions.

$$TTP_H = \int_{\xi_{low}}^{\xi_{cut}} \left[\frac{C_{TGT}}{CTFH_{sys}(\xi)} \right]^{1/2} d\xi \tag{8.46}$$

$$TTP_V = \int_{\eta_{low}}^{\eta_{cut}} \left[\frac{C_{TGT}}{CTFV_{sys}(\eta)} \right]^{1/2} d\eta \tag{8.47}$$

"Cycles on target" $N_{resolved}$ is found using Equation 8.48.

$$N_{resolved} = \frac{\sqrt{A_{TGT}\, TTP_H TTP_V}}{Range} \tag{8.48}$$

The out-of-band Spurious Response Ratio (SRR_{out}) is found for both horizontal and vertical, and $N_{resolved}$ is corrected for the presence of sampling artifacts; see Part 7.

$$N_{sampled} = N_{resolved} \sqrt{1 - 0.58\, SRRH_{out}} \sqrt{1 - 0.58\, SRRV_{out}} \tag{8.49}$$

The TTPF is used to find the probability of task performance.

$$P = \frac{\left(N_{sampled} / V_{50} \right)^E}{1 + \left(N_{sampled} / V_{50} \right)^E} \tag{8.50}$$

where

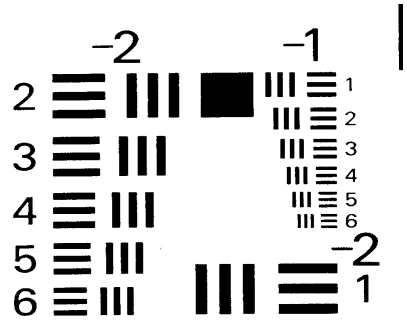
$$E = 1.51 + 0.24\, N_{sampled} / V_{50} \tag{8.51}$$

8.6 Minimum Resolvable Contrast

In the laboratory, sensors are characterized using Air Force 3-bar charts; a chart is shown in Figure 8.15. Each bar pattern is five times longer than the width of a single bar. Generally, charts with 1.0 contrast are used in the laboratory, but charts are available with lower contrast (but generally the contrast is above about 0.2). A plot of threshold contrast versus spatial frequency is called Minimum Resolvable Contrast (MRC). A plot of limiting frequency versus illumination level for a particular contrast is called a *limiting light* measurement.

When predicting the results of an MRC or limiting light experiment, the amplitude difference between the center bar and the adjoining spaces is used in place of the system MTF. The amplitudes are calculated as follows.

Figure 8.15 Air Force 3-bar chart used to characterize reflected-light imagers.



$$A_{center} = W \int_{-\infty}^{\infty} H_{sys}(\xi') H_W(\xi')[1 + 2\cos(4\pi W\xi')] d\xi' \tag{8.52}$$

$$A_{space} = W \int_{-\infty}^{\infty} H_{sys}(\xi') H_W(\xi')[1 + 2\cos(4\pi W\xi')]\cos(2\pi W\xi') d\xi' \tag{8.53}$$

$$S_L = L \int_{-\infty}^{\infty} H_{sys}(\xi') H_{eye}(\xi' / SMAG) H_L(\xi') d\xi' \tag{8.54}$$

where

ξ' = dummy variable for integration
W = 1/(2ξ)
L = 5W
$H_L(\xi)$, the bar-length MTF, is sin (πfL) / (πfL)
$H_W(\xi)$, the bar-width MTF, is sin (πfW) / (πfW)
S_L = Fractional intensity due to blur of bar length

The relationship between CTF_{sys} and MRC is:

$$MRC(\xi) = \left[\frac{H_{sys}(\xi) CTF(\xi)}{\left(A_{center}(\xi) - A_{space}(\xi)\right) S_L(\xi)} \right] \tag{8.55}$$

9

Modeling Thermal Imagers

Thermal imagers sense heat energy with wavelengths between three and twelve microns. The three to five micron band is called mid-wave IR (MWIR) and the eight to twelve micron band is called long-wave IR (LWIR). Figure 9.1 shows typical atmospheric transmission for a one Kilometer, horizontal path; there are three clear windows from 3 to 4.2, 4.4 to 5, and 8 to 13 microns.

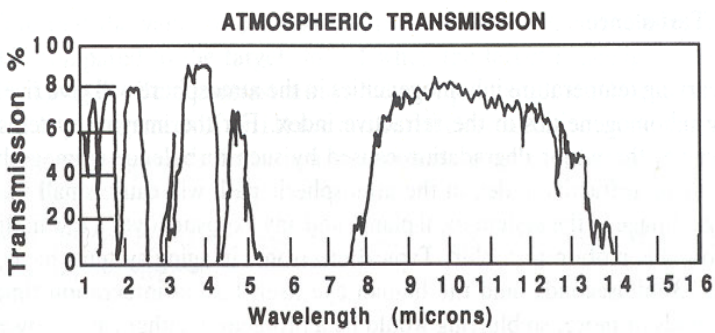


Figure 9.1 Atmospheric Transmission over 1 Kilometer Path

Figure 9.2 shows a schematic diagram of a thermal imager which uses a staring focal plane array (FPA) of detectors. The thermal scene is imaged by the objective lens onto the FPA. The individual detector signals are time multiplexed and converted to a video signal for display.

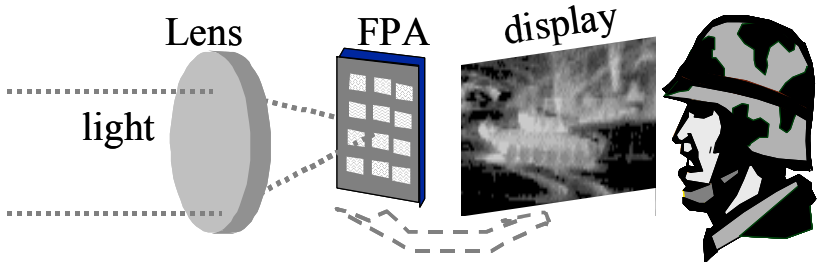


Figure 9.2 Illustration of Thermal Staring Sensor

Figure 9.3 shows a parallel scan thermal imager. The afocal provides a magnified image at the scanner. The scene is scanned over a linear array of detectors by an oscillating or rotating mirror. The time that each detector dwells on a point in the scene is less than that of the staring sensor; as a result, sensitivity is reduced.

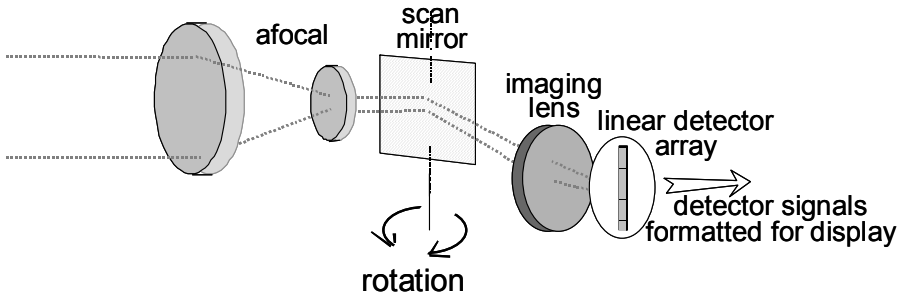
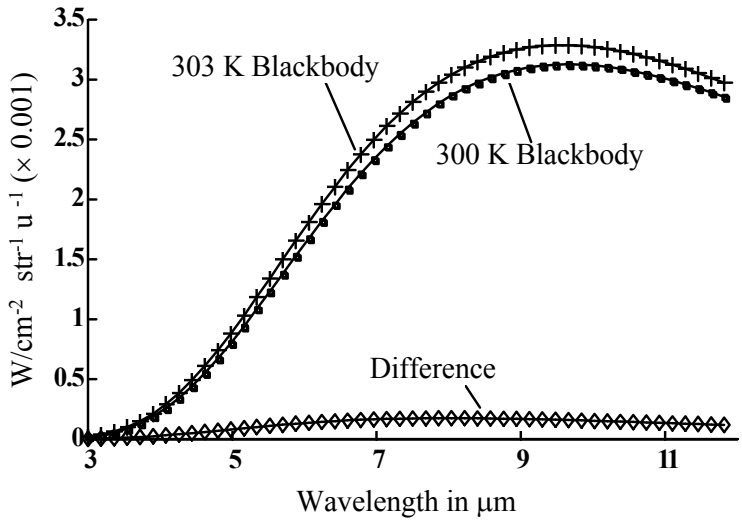


Figure 9.3 Illustration of Scanning Thermal Sensor

Everything near room temperature radiates at these wavelengths. The emissivity of natural objects is generally above 70 percent; most manmade objects are also highly emissive. Thermal sensors derive their images from small variations in temperature and emissivity within the scene. Typically, the thermal scene is very low contrast.

Figure 9.4 shows the spectral radiance from blackbodies at 300 K and 303 K. The difference between the two curves is also shown. As can be seen from the figure, a 3 K change in blackbody temperature results in only a small relative change in the radiated energy. However, a 3 K average for the apparent temperature difference within a scene represents very good thermal imaging conditions. A thermal imager will provide a good image under these conditions. The thermal scene is low contrast even under good thermal imaging conditions.

Figure 9.4 Thermal radiation from 300 K and 303 K blackbodies; both are near room temperature. Although the difference represents good thermal contrast, the relative difference is small.



Although the typical thermal scene is very low contrast, exceptions do exist. For example, the radiance difference between sky and ground can be quite large on a clear day. Also, the classic "burning tank" can overload a thermal imager. In general, however, thermal sensors are designed to map small differences in the scene's radiant energy into a usable displayed image.

In the above example, scene thermal contrast was generated by the temperature difference between two blackbodies. In the more general case, the spectral radiance from a thermal scene will depend upon a number of factors. The spectral radiance of an object

will depend upon its surface temperature and emissivity and upon the nature of the light being reflected or scattered from its surface. The apparent spectral radiance of an object as seen by an imager is also affected by the spectral transmission of the atmosphere. These factors, coupled with the spectral sensitivity of the imager itself, will determine the effective thermal contrast within the scene as sensed by a particular imager.

Apparent temperature (also called *equivalent blackbody temperature*) is often used as a radiometric unit. A radiometer is calibrated in terms of its response to a change in *blackbody temperature. The radiometer is then used to measure the thermal contrast of a scene, and its output is expressed as "temperature." The radiometer does not measure the temperature state of the scene; that is, the kinetic energy of the molecules in the scene objects is not measured. The radiometer is detecting the in-band energy from the scene, as weighted by the spectral response of the instrument itself. The effective blackbody temperature measured in one spectral band cannot be assumed for a different spectral band. When comparing MWIR to LWIR sensors, some knowledge is required of the relative signatures in the two spectral bands.

9.1 Signal and Noise in Thermal Imagers

The units used to describe signal and noise for thermal imagers are very different than the units used when modeling reflected-light sensors. However, aside from the details of calculating signal and noise, the basic CTF_{sys} theory is exactly the same as the theory described in Part 8.

The dominant noise in thermal photon detectors is generation recombination (GR) noise. In the theoretical limit, GR noise is the equivalent of the shot noise found in I^2 devices. However, noise can be increased by charge-carrier-phonon interactions. Thermal detectors are generally Background Limited in Performance (BLIP); noise decreases in proportion to the square root of detector photon flux. However, part of the background flux arises from within the imager itself, not just the scene. Even with perfect cold shielding, emission from the optics can be significant. Also, the read-out electronics adds noise, particularly with high F-number cold shields. Predicting the effect of reduced scene temperatures on noise is difficult. The noise from a thermal detector is very much dependent on system design and mounting factors as well as scene thermal flux. Generally, a thermal detector's noise characteristics are specified for a 300 K background temperature and a unique cold shield configuration.

Spectral detectivity ($D\lambda$) is used to specify the noise in a thermal detector.

$$D_\lambda = 1/NEP_\lambda \tag{9.1}$$

$NEP\lambda$ is the spectral noise-equivalent power; it is the monochromatic signal power necessary to produce an RMS signal to noise of unity. Spectral D-star ($D\lambda$*) is a normalization of $D\lambda$ to unit area and bandwidth.

$$D_\lambda * = D_\lambda \left(A_{det} \Delta f \right)^{1/2} \tag{9.2}$$

where

Δf = temporal bandwidth and
A_{det} = Active area of a single detector on the FPA = H_{det} V_{det}

The thermal model uses peak spectral D-star and relative detector response at other wavelengths to characterize detector performance.

$D^*\lambda_{peak} = D\lambda^*$ at wavelength of peak response and
$S(\lambda)$ = Response of detector at wavelength λ relative to peak response.

The spectral radiant power on the focal plane array is calculated as follows.

$$\boldsymbol{E_{fpa} = \pi \ \tau \ L_{scene} / 4 \ F\#^2} \tag{9.3}$$

where

E_{fpa} = watt cm^{-2} u^{-1} on the detector array,
L_{scene} = watts $cm^{-2}str^{-1}u^{-1}$ from the thermal scene, and
τ = transmission of optics.

The parameters τ, L_{scene}, and E_{fpa} are all functions of wavelength λ. The spectral radiant power on a single detector of the array is:

$$\boldsymbol{E_{det} = A_{det} \ \pi \ \tau \ L_{scene} / 4 \ F\#^2} \tag{9.4}$$

The signal to noise in one pixel (SN_{pix}) in one second can now be calculated.

$$SN_{pix} = \frac{D^*_{\lambda peak}}{(2A_{\det})^{1/2}} \int_{\Delta\lambda} E_{\det}(\lambda) S(\lambda) d\lambda \tag{9.5}$$

$$SN_{pix} = \left(\frac{D^*_{\lambda peak} (A_{\det})^{1/2} \pi\tau}{4\sqrt{2} F\#^2} \right) \left(\int_{\Delta\lambda} L_{scene}(\lambda) S(\lambda) d\lambda \right) \tag{9.6}$$

where $\Delta\lambda$ is the spectral band of the sensor, and 2 Hertz is the bandwidth Δf.

To estimate the differential spectral radiance resulting from a delta temperature near 300 K, the following equation is used. As long as the bars are at a temperature close to 300 K, the spectral nature of the difference signal is closely approximated by the partial derivative of the blackbody equation with respect to temperature evaluated at 300 K.

$$\boldsymbol{L_{\substack{scene \\ (Temp)}} = \Gamma \ \partial L(\lambda, T) / \partial T} \tag{9.7}$$

where the partial derivative is evaluated at 300 K and

$L(\lambda,T)$ = Plank's Equation for blackbody radiation,
T = Temperature, and
Γ = Amplitude of apparent blackbody temperature difference.

Using the Equation 9.7 expression for the spectral radiance based on temperature difference, the signal to noise on one detector in one second is now:

$$SN_{pix} = \frac{\Gamma\,\delta\,D^*_{\lambda peak}\left(A_{\det}\right)^{1/2}\pi\tau}{4\sqrt{2}\,F\#^2} \tag{9.8}$$

where

$$\delta = \int_{\Delta\lambda}\left(\partial \mathrm{L}(\lambda,\mathrm{T})/\partial \mathrm{T}\right)\mathrm{S}(\lambda)\,\mathrm{d}\lambda \tag{9.9}$$

In one square radian, the signal to noise would increase by an amount $(F_0{}^2/\mathrm{H_{pit}V_{pit}})^{1/2}$ where F_O is the effective focal length of the afocal or objective lens. The signal to detector noise in one second and one square milliradian is:

$$SN_{\det} = (1E-6)\,\Gamma\,\delta\,F_o\,\eta_{stare}D^*_{\lambda peak}\,\pi\tau\big/4\sqrt{2}F\#^2 \tag{9.10}$$

where η_{stare} is the square root of the fill factor ratio $\mathrm{H_{det}\,V_{det}/(H_{pit}\,V_{pit})}$. Equation 9.10 gives the signal to noise for temperature difference Γ. Noise modulation at the display is needed to find $\mathrm{CTF_{sys}}$. Setting signal to noise to unity, $\Gamma_{\det}$ is noise variance in units of (K-milliradian)^2.

$$\Gamma_{\det} = 4\sqrt{2}F\#^2\,/[(1E-6)\;\delta\,F_o\,\eta_{stare}D^*_{\lambda peak}\,\pi\tau] \tag{9.10}$$

9.2 $\mathrm{CTF_{sys}}$ for Thermal Imagers

Calculating $\mathrm{CTF_{sys}}$ requires that detector noise be expressed as display luminance noise. This in turn requires a mapping between radiometric temperature changes in the scene and the matching luminance changes on the display. The gain through the imager must be established in terms of foot-Lamberts per Kelvin. As with reflected-light imagers, the average and minimum display luminance is a model input.

Scene contrast temperature ($\mathrm{SCN_{TMP}}$) is the delta radiometric temperature in the scene needed to generate the average display luminance when minimum luminance is zero. Recall that the thermal image arises from small variations in temperature and emissivity within the scene, and these small variations are superimposed on a large background flux. Zero luminance on the display corresponds to the minimum scene radiant energy, not to zero radiant energy. SCN_{TMP} is not the absolute background radiometric temperature; it is the temperature contrast needed to raise the display luminance from zero to average. $\mathrm{SCN_{TMP}}$ is used rather than κ_{con} to indicate sensor gain state. A large $\mathrm{SCN_{TMP}}$ means gain is low, a small $\mathrm{SCN_{TMP}}$ means the gain is high.

With display noise modulation established, $\mathrm{CTF_{sys}}$ can be calculated.

$$CTFH_{sys}(\xi) = \frac{CTF\left(\xi\big/SMAG\right)}{M_{dsp}H_{sys}(\xi)}\left(1+\frac{\alpha^2\Gamma_{\det}\,QH_{hor}(\xi)\,QV_{hor}}{SCN_{TMP}^2}\right)^{1/2} \tag{9.11}$$

$$CTFV_{sys}(\eta) = \frac{CTF\left(\eta/SMAG\right)}{M_{dsp}V_{sys}(\xi)}\left(1 + \frac{\alpha^2\Gamma_{\det}\,QH_{ver}QV_{ver}(\eta)}{SCN_{TMP}^2}\right)^{1/2} \tag{9.12}$$

where

α = 169.6 (a proportionality constant)
ξ = horizontal spatial frequency in (milliradian)$^{-1}$
η = vertical spatial frequency in (milliradian)$^{-1}$
CTF(ξ/SMAG) = naked eye Contrast Threshold Function; see Appendix E
SCN_{TMP} = scene temperature which generates average display luminance
B(ξ or η) = the Equation (3.9) eye filters
H_{eye}(ξ or η) = eyeball MTF; see Appendix E
H_{elec}(ξ) = horizontal electronics MTF
V_{elec}(η) = vertical electronics MTF
H_{dsp}(ξ) = horizontal display MTF
V_{dsp}(η) = vertical display MTF
H_{sys}(ξ) = horizontal system MTF
V_{sys}(η) = vertical system MTF
QH_{hor} = horizontal noise bandwidth for $CTFH_{sys}$ defined by Equation 8.5
QV_{hor} = vertical noise bandwidth for $CTFH_{sys}$ defined by Equation 8.6
QH_{ver} = horizontal noise bandwidth for $CTFV_{sys}$ defined by Equation 8.7
QV_{ver} = vertical noise bandwidth for $CTFV_{sys}$ defined by Equation 8.8

$$QH_{hor}(\xi) = \int\left|B(\xi'/\xi)H_{elec}(\xi')H_{dsp}(\xi')H_{eye}\left(\xi'/SMAG\right)\right|^2 d\xi' \tag{9.13}$$

$$QV_{hor} = \int\left|V_{elec}(\eta)V_{dsp}(\eta)H_{eye}\left(\eta/SMAG\right)\right|^2 d\eta \tag{9.14}$$

$$QH_{ver} = \int\left|H_{elec}(\xi)H_{dsp}(\xi)H_{eye}\left(\xi/SMAG\right)\right|^2 d\xi \tag{9.15}$$

$$QV_{ver}(\eta) = \int\left|B(\eta'/\eta)V_{elec}(\eta')V_{dsp}(\eta')H_{eye}\left(\eta'/SMAG\right)\right|^2 d\eta' \tag{9.16}$$

η_{stare} is currently detector fill factor. However, due to limitations in photo-electron storage capacity, the FPA might not integrate signal for a full field time. The efficiency factor used in Equation 9.10 should be adjusted for detector integration time.

$$\eta_{stare} = t_{\text{int}}\,T_{CCD}\,H_{\det}\,V_{\det}/(H_{pit}\,V_{pit}) \tag{9.17}$$

where

t_{int} = detector integration time < = $1/T_{CCD}$
T_{CCD} = field rate (probably 60 Hertz)

H_{det} = horizontal active area of detector
V_{det} = Vertical active area of detector
H_{pit} = horizontal detector pitch
V_{pit} = vertical detector pitch

Aside from slightly different MTF considerations discussed below, the only change needed to model scanning imagers is to adjust the noise for the reduced dwell time. For scanning sensors, a different efficiency factor (η_{eff}) is used. The dwell time is reduced by the amount of detector area divided by the image area at the detector focal plane. Also, the scene is generally over-scanned by the scanner, either to look at thermal references or for turn-around, so η_{scan} is less than unity. For scanning thermal imagers, η_{eff} is used in Equation 9.10 rather than η_{stare}.

$$\eta_{eff} = \eta_{scan} N_{\det} H_{\det} V_{\det} / FOV_H FOV_V F_0^2 \tag{9.18}$$

where

η_{scan} = scan efficiency (generally 0.4 to 0.75),
N_{det} = total number of detectors, either in parallel or in Time Delay and Integrate,
Example: for a 240 by 4 in TDI FPA, N_{det} = 960
FOV_H = horizontal field of view of the imager in radians, and
FOV_V = vertical field of view of the imager in radians.

MTF of the optics, detector, and display are the primary contributors to system MTF for thermal imagers. Other likely sources of blur are jitter in the line-of-sight due to lack of stabilization, vibration of the display relative to the eye, and digital processing. For staring sensors, no MTF is associated with detector integration time. This is not true for scanning imagers, however. During photo-electron integration, the scene is scanned over the detector, so image blur results from temporal integration of the detector signal. This is an important source of blur in scanning imagers.

9.3 Predicting Probability versus Range

9.3.1 Contrast Transmission through the Atmosphere

Certain assumptions are made to simplify calculations. These assumptions constrain the scenario for which the model is appropriate.

a) The target and background are co-located; the target is viewed against local terrain. Range to the target and range to the background are the same.

b) Absorption as well as scattering in the atmosphere can be important. An interface to MODTRAN is provided. Since the spectral nature of the target-to-background signature is defined by Equation 9.9, this spectral weighting is included in the MODTRAN implementation.

c) Average radiance seen by the imager does not change with range. The background flux is from a 300K blackbody. Target-to-background signal

disappears into the average radiance. The 300 K radiance establishes sensor noise characteristics.

Apparent target contrast results from predicting the apparent radiometric temperature difference and then dividing by twice the scene contrast temperature. SCN_{TMP} is determined by hardware setup and the environment viewed. It is possible for those intimate with specific hardware designs to accurately determine SCN_{TMP}. However, some simplifying assumptions are sufficient for most modeling purposes.

a) When the imager is optimized on a specific target, SCN_{TMP} is between three and five times the target apparent RSS contrast. If an observer is attempting to identify a target, for example, it is reasonable to assume that the imager is optimized for the purpose. Since larger SCN_{TMP} results in lower contrast, an optimistic assumption is that SCN_{TMP} is three times larger than the target contrast.

b) When searching for a target, the imager gain is adjusted based on scene content and not changed. Thermal contrast between 0.1 and 0.3 K represents poor thermal scene contrast. Moderately good contrast is between 1 and 3 K; first generation thermal imagers (circa the mid 1980's) work well with thermal contrast in the 1 to 3 K range. Likely values for SCN_{TMP} are 1 K for poor thermal scenes and 5 to 10 K for good thermal contrast. However, when modeling search, do not input SCN_{TMP} less than three times the target intrinsic (zero range) thermal contrast. That is, if the target contrast is assumed to be 1.25 K, then SCN_{TMP} would be at least 3.75 K even if poor weather is assumed.

c) There are cases like thermal line-scanners where the total field of view is extremely wide. In these cases, SCN_{TMP} is likely to be 20 K or larger when thermal conditions are good.

d) Since SCN_{TMP} represents display average luminance, it is not physically possible for SCN_{TMP} to be less than half of the target to background thermal contrast. Given the realities of thermal signatures, SCN_{TMP} will realistically be three to five times the target RSS thermal signature as suggested above.

9.3.2 Effect of Contrast Enhancement

Contrast enhancement can significantly boost performance. As an example, an observer is searching using a LWIR imager on a fairly humid day with a 0.7 per kilometer transmission. If a 2 K target is at four kilometers, then apparent contrast is $(.7)^4$ 2/20 or about 0.024 contrast on the display. In this example, scene contrast temperature is taken as five times the target temperature or 10 K. It is assumed that the background scene content is fairly hot, and this really establishes the scene contrast temperature. If the observer detects the target and then optimizes the imager for target ID, SCN_{TMP} is adjusted (gain is increased) to a value five times the apparent temperature (SCN_{TMP} becomes 2.4 K), so that the contrast on the display is 0.1. Of course, the benefit of the improved contrast depends on the noise characteristics of the sensor, but the improved contrast could be significant.

The following "rule of thumb" is suggested.

a) When search is modeled, SCN_{TMP} is set based on scene contrast or at least three times the target signature, whichever is bigger. Atmospheric transmission reduces the apparent thermal signature (T_{TGT}) as range increases, and C_{TGT} is modeled as $C_{TGT} = T_{TGT}/(2\ SCN_{TMP})$.

b) When modeling target ID or any circumstance where contrast enhancement can be assumed, then C_{TGT} is fixed at the zero range value. If SCN_{TMP-0} is the input value (the zero range value) of scene contrast temperature, then $SCN_{TMP} = T_{TGT}\ SCN_{TMP-0} / T_{TGT-0}$.

9.3.3 Calculating Probability of Task Performance

At each range, apparent thermal contrast T_{TGT} is established based on zero range contrast T_{TGT-0} by using Beer's Law or MODTRAN. If no contrast enhancement is assumed, then $C_{TGT} = T_{TGT}/(2SCN_{TMP})$ and SCN_{TMP} remains constant at SCN_{TMP-0}. If contrast enhancement is used, then $C_{TGT} = T_{TGT-0}/(2SCN_{TMP-0})$ but the SCN_{TMP} used to calculate $CTFH_{sys}$ and $CTFV_{sys}$ decreases with range: $SCN_{TMP} = T_{TGT}\ SCN_{TMP-0} / T_{TGT-0}$. $CTFH_{sys}$ and $CTFV_{sys}$ are calculated and the TTP metric is found for both the horizontal and vertical dimensions.

$$TTP_H = \int_{\xi_{low}}^{\xi_{cut}} \left[\frac{C_{TGT}}{CTFH_{sys}(\xi)} \right]^{1/2} d\xi \tag{9.19}$$

$$TTP_V = \int_{\eta_{low}}^{\eta_{cut}} \left[\frac{C_{TGT}}{CTFV_{sys}(\eta)} \right]^{1/2} d\eta \tag{9.20}$$

"Cycles on target" $N_{resolved}$ is found using Equation 9.21.

$$N_{resolved} = \frac{\sqrt{A_{TGT}\ TTP_H TTP_V}}{Range} \tag{9.21}$$

The out-of-band Spurious Response Ratio (SRR_{out}) is found for both horizontal and vertical, and $N_{resolved}$ is corrected for the presence of sampling artifacts; see Part 7.

$$N_{sampled} = N_{resolved} \sqrt{1 - 0.58\, SRRH_{out}} \sqrt{1 - 0.58\, SRRV_{out}} \tag{9.22}$$

The TTPF is used to find the probability of task performance.

$$P = \frac{\left(N_{sampled} / V_{50} \right)^E}{1 + \left(N_{sampled} / V_{50} \right)^E} \tag{9.23}$$

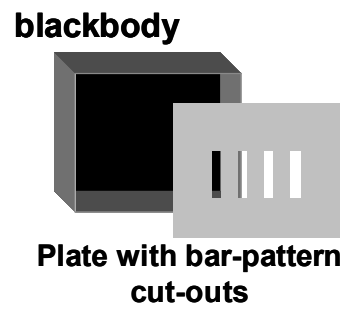
where

$$E = 1.51 + 0.24\, N_{sampled} / V_{50} \tag{9.24}$$

9.4 Minimum Resolvable Temperature

In the laboratory, thermal sensors are characterized using 4-bar patterns. The bar-target arrangement is shown in Figure 9.4. The bars are cut into a blackened, metal plate which is mounted in front of a blackbody. The temperature difference between the plate and blackbody is controlled. Each bar pattern is seven times longer than the width of a single bar. A plot of temperature difference between the bars and spaces versus spatial frequency is called Minimum Resolvable Temperature (MRT).

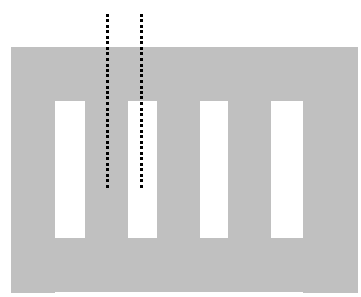
Figure 9.4 4-bar pattern used for MRT. Blackbody is viewed through the openings of metal plate.



MRT is a poorly controlled measurement. The imager gain and level are "optimized" for each bar size; saturation is permitted. Display luminance and contrast are not controlled or measured. The imager's settings are not monitored, and the bar targets are not viewed in a fashion that correlates to field performance. Experience over many years suggests that only a gross estimate of laboratory MRT can be predicted. MRT should only be used as an indicator that the system is operating at some acceptable level.

The bar pattern positions for which the modulation difference is calculated are shown by the dotted lines in Figure 9.5 This modulation difference is used in place of the system MTF. Assume that the white area is hotter and is called the bar. The amplitudes are calculated as follows.

Figure 9.4 4-bar pattern showing positions where difference modulation is calculated.



$$H_{4-bar} = W \int_{-\infty}^{\infty} H_{sys}(\xi') H_W(\xi')[2\cos(2\pi W\xi') + 2\cos(6\pi W\xi')]d\xi' \tag{9.25}$$

$$A_{bar} = W \int_{-\infty}^{\infty} H_{sys}(\xi') H_{4-bar}(\xi')\cos(2\pi W\xi')\, d\xi' \tag{9.26}$$

$$A_{space} = W \int_{-\infty}^{\infty} H_{sys}(\xi') H_{4-bar}(\xi')\cos(4\pi W\xi')\, d\xi' \tag{9.27}$$

$$S_L = L \int_{-\infty}^{\infty} H_{sys}(\xi') H_{eye}(\xi'/SMAG) H_L(\xi')\, d\xi' \tag{9.28}$$

where

ξ' = dummy variable for integration
W = 1/(2ξ)
L = 7W
$H_L(\xi)$, the bar-length MTF, is sin (πfL) / (πfL)
$H_W(\xi)$, the bar-width MTF, is sin (πfW) / (πfW)
S_L = Fractional intensity due to blur of bar length

The relationship between CTF_{sys} and MRT is:

$$MRT(\xi) = \left[\frac{H_{sys}(\xi)\, CTF_{sys}(\xi)}{\left(A_{bar}(\xi) - A_{space}(\xi)\right) S_L(\xi)} \right] \tag{9.29}$$

References:

Aguilar M and Stiles WS (1954), "Saturation of the rod mechanism of the retina at high levels of stimulation," *Optical Act.* 1: 59-65.

Barnes, R.B. Von, and M. Czerny (1933), "Can a Photon Shot Effect be observed with the Eye?", (in German), *Z. Physic*, Vol. 79, 436.

Barten, P.G.J., (1999), *Contrast Sensitivity of the Human Eye and its Effects on Image Quality,* SPIE Press, Bellingham, WA.

Beaton, R.J., and W.W. Farley (1991), "Comparative study of the MTFA, ICS, and SQRI image quality metrics for visual display systems," , Armstrong Lab., Air Force Systems Command, Wright-Patterson AFB, OH, Report AL-TR-1992-0001, DTIC ADA252116.

Bender, Edward (2000), "Present Image Intensifier Tube Structures," In *Electro-Optical Imaging: System Performance and Modeling*, Chapter 5, Lucien Biberman Ed., SPIE and ONTAR Corp., Bellingham, WA.

Biberman, Lucien (1973), "Image Quality," In *Perception of Displayed Information*," Chapter 2, Lucien Biberman, Ed., Plenum Press, N.Y.

Biberman, Lucien (1973), "Summary," In *Perception of Displayed Information*," Chapter 8, Lucien Biberman, Ed., Plenum Press, N.Y.

Biberman, Lucien (Editor) (2000), *Electro-Optical Imaging: System Performance and Modeling*, SPIE and ONTAR Corp., Bellingham, WA..

Biberman, Lucien (2000), "System performance and image quality," In *Electro-Optical Imaging: System Performance and Modeling*, Chapter 22, Lucien Biberman, Ed., SPIE and ONTAR Corp., Bellingham, WA.

Biberman, Lucien (2000), "Alternate modeling concepts," In *Electro-Optical Imaging: System Performance and Modeling*, Chapter 11, Lucien Biberman, Ed., SPIE and ONTAR Corp., Bellingham, WA.

Biederman, I. and E.E. Cooper (1987), Sexing Day-Old Chicks: A case study in expert systems analysis of a difficult perceptual learning task," J. Exp. Psyc.: Human Learning, Memory and Cognition, Vol. 13, 640-645.

Bijl, P. and J. M. Valeton, (1998), "TOD, the alternative to MRTD and MRC," OE 37(7), 1976-1983.

Blackwell, H.R. and O.T. Law (1958), U. MI Eng. Research Inst. Report 2455-9-F, cited in Overington (1976), Chapter 6, 115-116.

Boff, K. R. and J. E. Lincoln (1988), *Engineering Data Compendium Human Perception and Performance*, Vol. 2, Armstrong Laboratory, Wright-Patterson AFB, Ohio (1988).

Brock, G.C. (1965), Paper presented before the Soc. Photo. Sci. Eng. (May).

Burle Industries (1974), *Electro-Optics Handbook*, Previously RCA EO Handbook, TP-135, Tube Products Division Lancaster, PA.

Chen, J.S., Tsou, B.H., and Grigsby, S.S. (1994), "A Study on Contrast Perception in Noise," SID, Vol. XXV, 490-493.

Coltman, J. W. (1954), "Scintillation Limitations to Resolving Power in Imaging Devices," *J. Opt. Soc. Am.* 44(3) 234–237.

Davson, H. (1990) *Davson's Physiology of the Eye, 5th ed*. London: Macmillan Academic and Professional Ltd.

Devitt, Nicole, Ronald G. Driggers, Richard Vollmerhausen, (2001), Steve Moyer, Keith Krapels, and John O'Connor, "Target Recognition Performance as a Function of Sampling," Proc. SPIE, Vol. 4372, p. 74-84, Infrared Imaging System: Design, Analysis, Modeling, and Testing XII, Gerald C. Holst; Ed., Sept.

De Vries, Hessel, (1943), "The Quantum Character of light and its Bearing upon Threshold of Vision, the Differential Sensitivity and Visual Acuity of the Eye," Physica, vol. 10, No. 7, 553-564.

Doll, Theodore J., et. al., (1998), "Robust, sensor-independent target detection and recognition based on computational models of human vision," OE, Vol. 37, No. 7, 2006-2021, July.

Driggers, Ronald G., P. Cox, T. Edwards (1999), *Introduction to Infrared and Electro – Optical Systems*, Artech House, Boston.

Driggers, Ronald G.; Vollmerhausen, Richard H.; Krapels, Keith A., (2000), "Target identification performance as a function of low spatial frequency image content," *Optical Engineering* 39(09), 2458-2462.

Driggers, R.G., R. Vollmerhausen, W. Wittenstein, P. Bijl, M. Valeton, (2000), "Infrared Imager Models for Undersampled Imaging Systems," Proceedings of the Fourth Joint Military Sensing Symposium, vol. 45, pp 335-346.

Green D. G. (1970) Regional variations in the visual acuity for interference fringes on the retina. *J Physiol.* Vol. 207:351-356.

Greis, U. and R. Rohler (1970), "Untersuchung der subjektiven Detailerkennbarkeit mit Hilfe der Ortsfrequenzfiltilterung," Optica Acta, Vol. 17, 515-526. (A translation by Ilze Mueller with amendments by D.G. Pelli, "A study of the subjective detectability of patterns by means of spatial-frequency filtering," is available from D.G. Pelli.)

Howe, James (1993), "Electro-Optical Imaging System Performance Prediction," In Electro-Optical Systems Design, Analysis, and Testing, (Dudzik Ed.), The IR & EO systems Handbook, Vol 4, pg. 92, ERIM IIAC, Ann Arbor, MI and SPIE, Bellingham, WA.

Johnson, J., (1958), "Analysis of imaging forming systems," *Proceedings of the Image Intensifier Symposium,* October 6-7, 1958, AD220-160, U.S. Army Engineer Research and Development Lab, Fort Belvoir, VA, 249-273.

Kornfeld, G.H., and W.R. Lawson (1971), "Visual Perception Models," JOSA, Vol. 61, No. 6, 811-820.

Krapels, K., R. Driggers, R. Vollmerhausen, C. Halford, (1999), "The Influence of Sampling on Recognition and Identification Performance," *Optical Engineering* Vol. 36, No. 5, May.

Krapels, D., R. Driggers, and R. Vollmerhausen, (2001),"Performance Comparison of Rectangular (Four Point) and Diagonal (Two Point) Dither in Undersampled Infrared Focal Plane Array Imagers," *Applied Optics*, Vol 40, No 1, Jan., pg 71-84.

Lawson, Walter R. (1971), Electrooptical System Evaluation, In *Photoelectronic Imaging Devices*, Vol.1, pg. 375, L. Biberman and S. Nudelman, eds., Plenum Press, N.Y.

Lawson, W.R., and J.A. Ratches (1979), "The Night Vision Laboratory Static Performance Model based on the Matched Filter Concept," In "The Fundamentals of Thermal Imaging Systems," by Fred Rosell and George Harvey, EO Technology Program Office, NRL, Washington, D.C., DTIC ADA073763.

Legge, G.E., et. al.(1987), "Contrast discrimination in noise," J. Opt. Soc. Am., Vol 4., No. 2, 391-404.

Lu, Zhong-Lin and Barbara Anne Dosher (1999), "Characterizing human perceptual inefficiencies with equivalent internal noise," JOSA A, Vol.16, No. 3.

Nagaraja, N.S. (1964), "Effect of Luminance Noise on Contrast Thresholds," J. Opt. Soc. Am., Vol 54, No. 7, 950-955.

Normann, R.A., Perlman, I. and Hallet, P.E. (1991) Cone photoreceptor physiology and cone contributions to colour vision. In "Vision and Visual Dysfunction volume 6", "The Perception of Colour". (Ed. Gouras, P.) Macmillan Press Ltd., London, 146-162.

O'Kane, Barbara, Irving Biederman, and Eric Cooper (2000), "Modeling parameters for target identification: A critical features analysis," In *Electro-Optical Imaging: System Performance and Modeling*, Chapter 15, Lucien Biberman Ed., SPIE and ONTAR Corp., Bellingham, WA.

Osterberg, G. (1935) Topography of the layer of rods and cones in the human retina. Acta Ophthal., suppl. **6**, 1-103.

Overington, Ian (1976), *Vision and Acquisition*, Crane, Russak & Company, N.Y.

Pelli, D.G., (1981), "Effects of visual noise," Doctoral dissertation at the Physiological Laboratory, Churchill College, Cambridge University, England. Available in PDF from denis.pelli@nyu.edu.

Pelli, D.G., (1999), "Why use noise?" JOSA A, Vol. 16, No. 3, March 1999.

Raghavan, M. (1989), "Sources of visual noise," Ph.D. dissertation (Syracuse Univ., Syracuse, New York.

Ratches, J. A., et al. (1975), "Night Vision Laboratory Static Performance Model for Thermal Viewing Systems," Research and Development Technical Report ECOM–7043, U. S. Army Electronics Command, Fort Monmouth, New Jersey.

Ratches, J. A., (1976), "Static Performance Model for Thermal Imaging Systems," OE, Vol. 15, No. 6, Nov.-Dec.

Ratches, J. A., Richard Vollmerhausen, Ron Driggers (2001), Target Acquisition Performance Modeling of Infrared Imaging Systems: Past, Present, and Future," IEEE Sensors Journal, Vol. 1, No. 1, 31-40, June.

Richards, E. A. (1967), "Fundamental Limitations in the Low Light-Level Performance of Direct-View Image-Intensifier Systems," *Infrared Physics*, Pergamon, Great Britain.

Rose, Albert (1948), "The Sensitivity Performance of the Human Eye on an Absolute Scale," *J. Opt. Soc. Am.* 38(2):196–208.

Rosell, Fred and R.H. Wilson (1973), "Recent Psychophysical Experiments and the Display signal to Noise Ratio Concept," In *Perception of Displayed Information*," Chapter 5, Lucien Biberman, Ed., Plenum Press, N.Y.

Rosell, Fred, and George Harvey (1979), "The Fundamentals of Thermal Imaging Systems," EO Technology Program Office, NRL Report 8311, Washington, D.C., DTIC ADA073763.

Rosell, Fred and R.L. Sendall (2000), "Static Performance Model based on the Perfect Synchronous Integrator Model," Chapter 13 in *Electro-Optical Imaging: System Performance and Modeling*, Lucien Biberman Ed.,SPIE and ONTAR Corp.

Rosell, Fred (2000), "Synthesis and Analysis of Imaging Sensors," Chapter 14 in *Electro-Optical Imaging: System Performance and Modeling*, Lucien Biberman Ed.,SPIE and ONTAR Corp.

Schade, Otto Sr., (1948), "Electro-Optical Characteristics of Television Systems," Parts I-IV, RCA Rev., Vol. 9, March: 5-37, June: 247-286, Sept.: 491-531, Dec.: 653-687.

Schade, Otto Sr. (1956), "Optical and Photoelectric Analog of the Eye," JOSA, Vol. 49, No. 9, 721-739.

Schade, Otto Sr. (1973), "Image Reproduction by a Line Raster Process," In *Perception of Displayed Information*," Chapter 6, Lucien Biberman, Ed., Plenum Press, N.Y.

Schnitzler, Alvin D. (1973), "Analysis of Noise Required Contrast and Modulation of Image-Detecting and Display Systems," In *Perception of Displayed Information*," Chapter 4, Lucien Biberman, Ed., Plenum Press, N.Y.

Sendall, R.L. and Fred Rosell (1979), "Static Performance Model based on the Perfect Synchronous Integrator Model," In "The Fundamentals of Thermal Imaging Systems," by Fred Rosell and George Harvey, EO Technology Program Office, NRL, Washington, D.C., DTIC ADA073763.

Snyder, Harry L. (1973), Image quality and observer performance," In *Perception of Displayed Information*," Chapter 3, Lucien Biberman, Ed., Plenum Press, N.Y.

Snyder, Harry L. (1985), "Image Quality: Measures and Visual Performance," In *Flat Panel Displays and CRTs*, Ch. 4, L.E. Tannas, Ed., Van Nostrand Reinhold Company, New York.

Snyder, Harry L. (1988), "Image Quality," In *Handbook of Human-Computer Interaction*, Chapter 20, M. Helander Ed., Elesier Science Publishers B.V., North-Holland.

Stromeyer, C.F., and B. Julesz (1972), "Spatial frequency masking in vision: critical bands and spread of masking," JOSA, 62, 1221-1232.

Tannas, L. E., (1985), *Flat Panel Displays and CRTs,* Ch.4, Van Nostrand Reinhold Company, New York.

Task, H. L., (1976), "An Evaluation and Comparison of Several Measures of Image Quality for Television Displays," Technical Report AMRL–TR–76–73, Air Force Aerospace Medical Research Laboratory, Wright-Patterson AFB, Ohio.

Van Meeteren, A. & Valeton, J. (1988), "Effects of pictorial noise interfering with visual detection," JOSA A, Vol 5, No. 3, 438-444.

van Meetern, A. (1990), Characterization of task performance with viewing instruments," JOSA A, Vol. 7, No. 10, 2016-2023, October.

Vollmerhausen, R. (1995), "Incorporating Display Limitations into Night Vision Performance Models," *1995 IRIS Passive Sensors*, V2**:**11-31.

Vollmerhausen, R. (1996), "Minimum Resolvable Contrast Model for Image Intensified Charge Coupled Device Cameras," U. S. Army CECOM Night Vision and Electronic Sensors Directorate, Technical Report NV–96–09.

Vollmerhausen, R., R. Driggers, and B. O'Kane, (1999), "Character Recognition as a Function of Spurious Response," *Journal of the Optical Society of America A*, Vol. 16, No 5, May.

Vollmerhausen, Richard H. (2000), "Modeling the Performance of Imaging Sensors," In *Electro-Optical Imaging: System Performance and Modeling*, Chapter 12, Lucien Biberman Ed., SPIE and ONTAR Corp., Bellingham, WA.

Vollmerhausen, Richard H. (2000), "Display of Sampled Imagery," In *Electro-Optical Imaging: System Performance and Modeling*, Chapter 25, Lucien Biberman Ed., SPIE and ONTAR Corp., Bellingham, WA.

Vollmerhausen,. Richard H. and Ronald G. Driggers, (04/2000), *Sampled Imaging Systems*, SPIE Tutorial Series.

Vollmerhausen, Richard H.; Driggers, Ronald G.; Tomkinson, Michelle, (07/2000), "Improved image quality metric for predicting tactical vehicle identification," *Proc. SPIE Vol. 4030,* p. 60-69*, Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XI,* Gerald C. Holst; Ed.

Vollmerhausen, Richard H., Eddie Jacobs, and Ronald Driggers (2003), "New metric for predicting target acquisition performance," Proc. SPIE Vol. 5076, *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XIV,* Gerald C. Holst; Ed.

Vos, Johannes, and Aart van Meeteren (1991), "PHIND: an analytic model to predict target acquisition distance with image intensifiers," Applied Optics, Vol. 30, No. 8, 958-966, March.

Webvision (2003) , "The organization of the retina and visual system," edited by Eduardo Fernandez, Ralph Nelson, Helga Kolb, and Bryan William Jones, on the web at Webvision.med.utah.edu, Part IX, "Psychophysics of Vision," by Michael Kalloniatis and Charles Luu , Sections 2 and 4, September, 2003.

Wittenstein, W., (1999), "Minimum temperature difference perceived--a new approach to assess undersampled thermal imagers" AA (FGAN-Forschungsinstitut fuer Optik) Publication: Optical Engineering 38(05), 773-781, Donald C. O'Shea; Ed.

Yau, K.-W. (1994) Phototransduction mechanisms in retinal rods and cones. Invest. Ophthal. Vis. Sci. **35**, 9-32.

Appendix A: Description of Validation Experiments

A robust performance metric must provide accurate predictions for various shape and size blurs, good and poor intrinsic target contrast, various levels and types of noise, and must accurately predict the performance of sampled imagers. A series of experiments was designed to ensure the metric is accurate for a wide variety of image characteristics. Some of the experiments are not reported here. Experiments which demonstrated that the model works with both fixed spatial noise and temporally random noise are reported in Reference 23. The model is applicable to both framing and snapshot imagers. Experiments with white noise which had both uniform and Gaussian amplitude distributions are also reported in that reference. We found that only the RMS noise level mattered; the nature of the amplitude distribution did not matter. This paper describes experiments to test the TTP metric with various types and levels of blur, noise, and contrast. Experiments with sampled imagery are also reported in this paper.

Two displays were used in these experiments. The color display was a high quality CRT, computer monitor. On this display, a 200 pixel image measured 3 inches. The color display was operated in a mode with 8 bit quantization of the video. Typically, subjects viewed this display from about 18 inches; however, the viewing distance was not constrained with a chin rest or bit-bar. A high resolution, black and white display was also used. This display provided 10 bit quantization of the output video. On this display, 591 pixels measured 3.25 inches. Typically, subjects viewed the black and white display from about 15 inches. The viewing distance was not constrained except in Experiment 25. In all of the experiments, the average display luminance was five fL. Gamma correction was not used for either display. The estimated MTF for the color display and the measured MTF for the black/white display are shown in Table A1.

Table A1 Display MTF

Spatial frequency cycles/milliradian	0.1	0.2	0.3	0.4	0.5	0.6	0.8	1.0	1.2	1.4
Black/white horizontal MTF	.98	.94	.86	.77	.66	.56	.35	.20	.10	.04
Black/white vertical MTF	.98	.92	.83	.71	.59	.47	.26	.12	.05	.02
Color horizontal MTF	.96	.85	.69	.52	.36	.22	.06	.01		
Color vertical MTF	.97	.89	.77	.63	.48	.35	.15	.05	.02	

A.1 Target Acquisition Task

All of the experiments reported here involved target identification (ID). In these experiments, the observers were trained to identify twelve targets. Images of the targets were then degraded by blurring, adding noise, or reducing contrast, and the observers were asked to identify the target based on the degraded image.

The twelve targets are shown in Figures A1 and A2. Both thermal images and visible images were used; only examples of the thermal imagery are shown here. Figure A1 shows only a side aspect of each target, but twelve aspects of each of the twelve targets were used during the experiments. Figure A2 shows these aspects for the T55 Russian tank. Pristine image sets were collected in both the thermal and visible spectral bands. The size of the images was 401 by 590 pixels. The square root of target area, averaged over all aspect angles, is 3.11 meters. The target range during imagery collection was 125 meters.

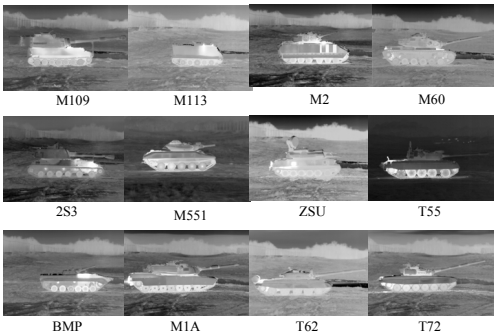


Figure A1. Illustration of the 12 targets used for experiments.

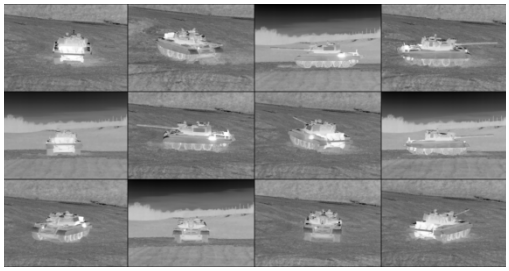


Figure A2. Illustration of the 12 aspects used for each target shown in Figure A1. Target shown is a T55 Russian tank.

These images were processed to generate the experiments. In all of the experiments, a cell of 24 images was created for each combination of MTF, noise, contrast, or range. Each cell contained two aspects of all twelve targets. Each of the aspects shown in Figure 6 was represented twice in each cell. This aspect distribution was chosen to make task difficulty as uniform as possible between cells. With 144 total images and 24 images per cell, up to six blur and noise combinations were created without repeating the use of an original image; these six cells created a "line" in the experiment. In all of the experiments, comparisons between different MTF types or different noise or contrast levels used the same original images to create the cells to be compared. This means that the same image was viewed three to five times in one experiment. Since the experiments contained between 432 and 720 images, it was doubtful that subject learning occurred because of repeating the images. Cell presentation was random based on a pre-selected order; all subjects saw the images in the same order. All of the images in one cell were presented sequentially.

The subjects were all active military and experienced in the use of thermal imagers. In addition, each observer was trained to ID the tactical vehicles used in these tests. All observers passed a pre-test with at least 95% correct; most observers passed with a 100% score. The average number of subjects for an individual experiment was 15 but varied from a minimum of 9 to a maximum of 23. The perception laboratory was moved between Army bases in order to maintain a large subject pool.

A.2 Description of Experiments

A.2.1 Experiments with well-sampled imagery

Gaussian (G), exponential (E), rect, and Difference of Gaussian (DOG) MTF were applied to the images; these MTF types are illustrated in Figure A3. Noise levels varied from zero to levels that completely obscured the targets. The magnification varied, depending on the experiment, from 0.63 to 2.8. Experiments were also performed with average target contrast ranging from 0.013 to 0.205. In some cases the images were down-sampled, noise added, and then the images were "interpolated up" or electronically zoomed. This process was needed to increase the impact of the noise on performance. Table A2 provides details for the MTF and white-noise experiments. In the table, the $e^{-\pi}$ (0.043) amplitude is used to define frequency cutoff. Frequency is in units of cycles per milliradian in object space. RMS noise is in units of fL. Noise was random with a Gaussian amplitude distribution. Noise was added to the imagery at run time; one of 240 pre-stored noise frames was randomly selected and added to the target image at the 60 Hertz display rate. None of the imagery in these experiments exhibited sampling artifacts. Pre- and post-filtering was always sufficient that no sampling artifacts were visible.

Figure A.3 Various types of MTF are shown. The dotted line indicates the 0.05 value used to indicate frequency cutoff in the table. Frequency in cycles per milliradian at the eye.

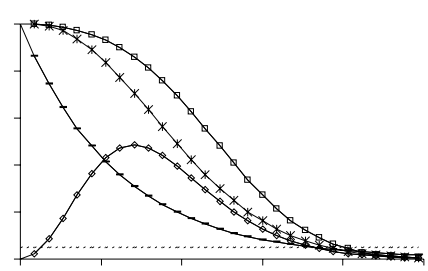


Table A2. MTF, noise, and contrast experiments

Experiment	MTF Type	MTF 0.043 cutoff object space cy/millirad	Contrast	Down-sample	RMS noise	zoom	Display & Magnification
6a line 1	G	.11,.14,.17,.23,.34,.68	0.205	By 2	0.0	No	Color (2.85)
6a line 2	E	.18,.21,.27,.36,.54,1.1	0.205	By 2	0.0	No	Color (2.85)
6b line 1	G	.11,.14,.17,.23,.34,.68	0.205	By 2	0.0	No	Color (2.85)
6b line 2	E	.11,.14,.17,.23,.34,.69	0.205	By 2	0.0	No	Color (2.85)
6b line 3	E	.1,.12,.14,.19,.29,.57	0.205	By 2	0.0	No	Color (2.85)
9 line 1	G	.18,.21,.26,.35,.53,1.05	0.205	By 2	0.0	No	Color (2.85)
9 line 2	E	.07,.08,.1,.13,.2,.41	0.205	By 2	0.0	No	Color (2.85)
9 line 3	DOG	.14,.19,.23,.31,.46,.91	0.205	By 2	0.0	No	Color (2.85)

9 line 4	Rect	.23,.28,.34,.46,.71,1.4	0.205	By 2	0.0	No	Color (2.85)
13 line 1	G	.23,.27,.34,.46,.68,1.4	0.205	By 8	0	By 2	Color (1.42)
13 line 2	G	.23,.27,.34,.46,.68,1.4	0.205	By 8	0.37	By 2	Color (1.42)
13 line 3	G	.23,.27,.34,.46,.68,1.4	0.205	By 8	0.74	By 2	Color (1.42)
13 line 4	G	.23,.27,.34,.46,.68,1.4	0.205	By 8	1.1	By 2	Color (1.42)
19 line 1	E	.36,.43,.54,.72,1.1,2.2	0.205	By 8	0	By 2	Color (1.42)
19 line 2	E	.36,.43,.54,.72,1.1,2.2	0.205	By 8	1.1	By 2	Color (1.42)
19 line 3	E	.36,.43,.54,.72,1.1,2.2	0.205	By 8	1.83	By 2	Color (1.42)
19 line 4	E	.36,.43,.54,.72,1.1,2.2	0.205	By 8	2.44	By 2	Color (1.42)
20 line 1	G	.23,.27,.34,.46,.68,1.4	0.205	By 8	0	By 2	Color (1.42)
20 line 2	G	.23,.27,.34,.46,.68,1.4	0.205	By 8	1.1	By 2	Color (1.42)
20 line 3	G	.23,.27,.34,.46,.68,1.4	0.205	By 8	1.83	By 2	Color (1.42)
20 line 4	G	.23,.27,.34,.46,.68,1.4	0.205	By 8	2.44	By 2	Color (1.42)
33 line 1	G	.2,.23,.27,.34,.46,.69	0.11	By 4	0	No	Mono (.63)
33 line 2	G	.2,.23,.27,.34,.46,.69	0.055	By 4	0	No	Mono (.63)
33 line 3	G	.2,.23,.27,.34,.46,.69	0.036	By 4	0	No	Mono (.63)
33 line 4	G	.2,.23,.27,.34,.46,.69	0.02	By 4	0	No	Mono (.63)
38 line 1	G	.42,.45,.48,.51,.55,.61	0.205	By 4	0	No	Mono (.63)

A.2.2 Experiments with sampled imagery

Sampling experiments were performed to show that the new metric works when a half-sample cutoff is imposed. That is, the TTP metric bases image quality on the un-corrupted frequency spectrum. Current models using the Johnson criteria cannot impose the half-sample cutoff, since this results in pessimistic performance predictions. In the sampling experiments, the blur, sampling, and display size were varied to represent the effect of increasing range.

Two experiments were performed. Experiment 25 examined the impact on range performance of different display interpolations. Visible display-pixel structure, like line raster or the edges of square pixels, tends to mask the underlying image and decrease range performance. Visible pixel structure is minimized by a good display interpolation which filters out spectrum beyond the half-sample frequency. In Experiment 25, aliased signal at less than the half-sample frequency was minimized. Experiment 36 was performed to explore the impact of large amounts of in-band aliasing on targeting performance. A small detector fill-factor was used to generate aliased signal at frequencies less than the half-sample rate.

The sensor simulated in Experiment 25 had the following characteristics. The mid-wave IR, staring focal plane array had 256 by 256 detectors. The active detector area was 28 microns on a 30 micron pitch. The sensor field of view was 2 by 2 degrees. The F/2 optics had a 22 centimeter focal length. The simulated ranges were 0.54 to 3.24 kilometers in 0.54 kilometer increments.

The imagery was displayed on the black and white monitor. Experiment 25 consisted of six lines each with six ranges (cells) with 24 target calls per cell. Each line used different interpolations to increase image size (electronically zoom the image); this changed the character is the displayed image by adding variable amounts of pixel structure. The display interpolations for each line are shown in Table A3. The kernel shown in Equation (A-1) provided the least amount of display structure; this kernel provides a good filter at the half-sample rate.

$$\text{Kernel} = [.011 \quad 0 \quad -.089 \quad 0 \quad .58 \quad 0 \quad .58 \quad 0 \quad -.089 \quad 0 \quad .011] \qquad \text{(A-1)}$$

Experiment 36 was performed to explore the impact of large amounts of in-band aliasing on targeting performance. Again, a 256 by 256 focal plane array was used. In this experiment, the detector pitch was 25 microns. The F/2 optics had a 7.33 centimeter focal length. Imagery was displayed on the black and white monitor. Simulated ranges were 0.43, 0.64, 0.97, 1.3, 1.6, and 2.15 kilometers.

Various amounts and types of aliasing were created by changing detector active area (detector fill factor) and display technique. In-band aliasing was varied by changing the detector fill factor. Low in-band aliasing resulted from setting the detector active area to 25 microns (100% fill factor). High in-band aliasing resulted from setting the detective active area to 1 micron (fill factor of 1/25 in both directions). Because the small detector fill-factor was associated with a high MTF, the MTF of the sensor used to collect the pristine images was significant and was modeled in this experiment.

$$\text{Sensor MTF} = e^{-0.82\, frequency^{1.2}} \qquad \text{(A-2)}$$

where frequency is in object space and has units of cycles per milliradian.

To change out-of-band aliasing (visibility of pixels), different display interpolations were used in Experiment 36 also; these are shown in Table A3. In all cases, sensor imagery was e-zoomed by 11 in both horizontal and vertical. Low out-of-band aliasing resulted from using the MATLAB bicubic image resize function to resize by eleven. The bicubic interpolation filtered out-of-band aliasing; no raster or pixel effects were visible. High out-of-band aliasing was created by using pixel replicate to e-zoom by eleven. In this case, the pixels were readily visible. The experiment consisted of four lines: (1) no in-band and no out-of-band aliasing, (2) no in-band with out-of-band, (3) in-band but no out-of-band, and (4) both in-band and out-of-band aliasing. The decrease in Nresolved due to sampling artifacts is shown in Table A3 as "sampling factor."

Table A3 Display interpolations for sampling experiment.

Experiment & line	1st interpolate	2nd interpolate	3rd interpolate	Total E-zoom	System magnification	Detector fill-factor	Sampling factor
25 line 1	None	Replicate	Replicate	4	9	Large	0.9
25 line 2	None	Bilinear	Replicate	4	9	Large	0.94
25 line 3	None	Equation (23)	Replicate	4	9	Large	0.97
25 line 4	Replicate	Replicate	Replicate	8	18	Large	0.82
25 line 5	Bilinear	Bilinear	Replicate	8	18	Large	0.93
25 line 6	Equation (23)	Equation (23)	Replicate	8	18	Large	0.97
36 line 1	None	None	Bicubic	11	10.6	Large	0.96
36 line 2	None	None	Replicate	11	10.6	Large	0.8
36 line 3	None	None	Bicubic	11	10.6	Small	0.73
36 line 4	None	None	Replicate	11	10.6	Small	0.62

Viewing distance is a problem with sampling experiments. In terms of performance degradation, the most serious type of sampling artifact is visible pixel structure. That is, display raster, pixels edges, and other periodic modulation beyond the half-sample frequency prevents the visual system from integrating the underlying picture. But eye MTF is a significant post filter. When sampling artifacts are present, the image can generally be seen better by moving the eye position away from the display. As viewing distance increases, eye MTF filters out the sampling artifacts. This behavior ruins the experiment. Unfortunately, in our facility, there is only one display station where viewing distance can be strictly controlled. This station was used for Experiment 25. The subjects were seated in a reclined chair such that the display viewing distance could be controlled at 18 inches. Viewing distance was not as well controlled for Experiment 36. The subjects were placed in fixed–back chairs without coasters and warned about excessive head movement. However, the subjects were not continually challenged to maintain head position. We did observe subjects sometimes lean back in an apparent effort to better discern the image.

A.2.3 Experiments with Colored Noise

Two experiments with colored noise were performed; the experiments were identical except that one used thermal images and the other used visible images. The contrast of the visible image set was 0.37; the contrast of the thermal set was 0.205. Each experiment consisted of four lines of six blurs each. The six blurs were created with a Gaussian kernel with $e^{-\pi}$ object-space frequency cutoffs of 0.2, 0.23, 0.27, 0.34, 0.46, and 0.69 cycles per milliradian. Magnification was 0.63, so that frequency cutoffs at the eye were 0.32, 0.37, 0.43, 0.54, 0.73, and 1.1. The images were blurred and then down-sampled by four. Frames of static, white noise were filtered and then added to the down-sampled images for display on the black and white monitors. The first line had no noise added, the second line had white noise added, the third line had low frequency noise added, and the forth line had high frequency noise added. The MTF of the noise filters

are shown in Figure A4. The RMS of the white noise before filtering was 0.98 fL for the white and high frequency noise lines. Before filtering, the RMS of the low frequency noise was 18 fL.

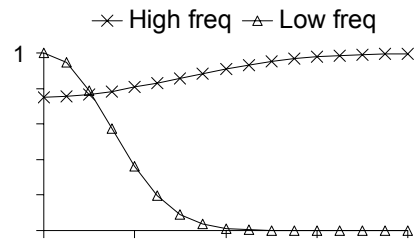


Figure A4. MTF of filters used to color noise. Spatial frequency is in object space.

A.3 Experimental Results

A.3.1 Well-Sampled Imagery

The results of the MTF, noise, and contrast experiments described in Section A.2.1 are shown in Figure A5 for the Johnson criteria and in Figure A6 for the TTP metric. In Figure A6, the abscissa is $N_{resolved}$ based on Equation (_). TTP values are calculated using Equation (_). When calculating the model predictions for Figure A5, the Johnson frequency (F_J) is found by the intersection of target contrast with the CTF_{sys} function. Equation (_) is used to find $N_{JCresolved}$ for the Johnson criteria also, but F_J is used rather than the TTP value. In both figures, the ordinate is probability of ID. In these figures, Experiments 6 and 9 data with various MTF shapes are designated with a diamond symbol (◊). Experiment 13 with Gaussian blur and low amounts of noise is designated by a square (□). Gaussian blur with large amounts of noise is designated by a triangle (∆); these data are from Experiment 20. Experiment 19 data representing exponential MTF with large amounts of noise are represented by asterisks (∗). Gaussian blur on the high resolution display, Experiment 38, is shown by open circles (❍). The low contrast experiment data from Experiment 33 are shown by filled-in circles (●).

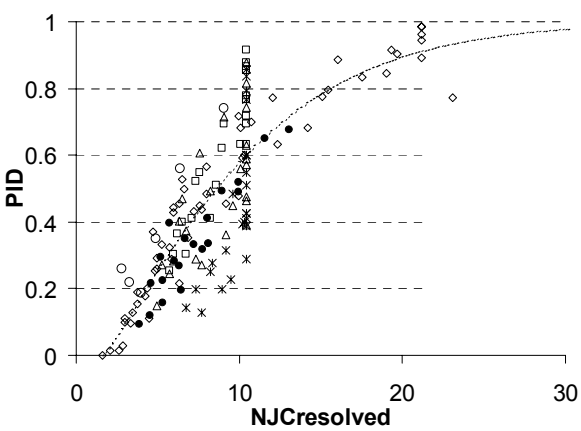


Figure A5. Results of MTF, contrast, and noise experiments for the Johnson criteria.

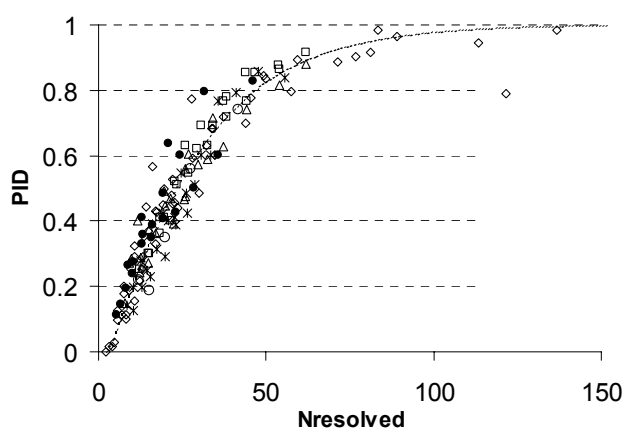


Figure A6. Results of MTF, contrast, and noise experiments for TTP metric.

The model curves shown in each figure are the Target Transfer Probability Function (TTPF) to use with each metric. The TTPF curves are logistics functions as defined by Equation (A-3). For the Johnson metric, N_{JC50} is 6.98. For the TTP metric, N_{50} is 20.8.

$$P_{ID} = \frac{\left(N/N_{50}\right)^E}{1+\left(N/N_{50}\right)^E} \tag{A-3}$$

where for the TTP metric

$$N = N_{resolved} - 2.6$$
$$E = 1.33 + 0.33\, N/N50 \tag{A-4}$$

and for the Johnson criteria

$$N = N_{JCresolved} - 1.6$$
$$E = 1.33 + 0.23\, N/JCN50 \quad . \tag{A-5}$$

The P_{ID} data represent the average number of correct calls for all observers for each cell of 24 target images. The experimental data are corrected in two ways. First, the probability of chance is taken out of the data. That is, the data are adjusted to remove the one in twelve chance that the subject will correctly ID the target by accident. The data are also corrected for mistakes. Experimentally, the ID probabilities asymptote to 0.9 rather than 1; there is a 10% mistake rate that does not correlate to cycles on target. Equation (A-6) is used to correct the measured data.

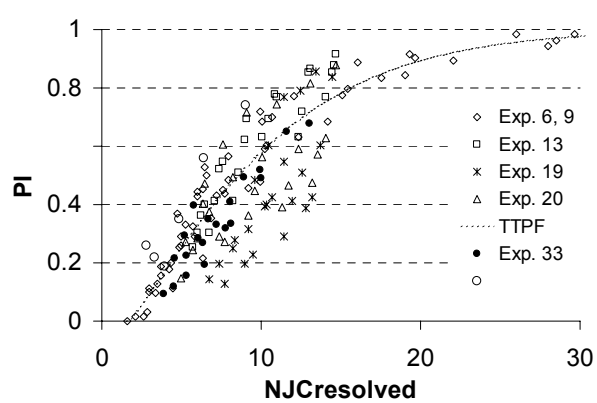
$$P_{ID} = \frac{P_{measured} - P_{chance}}{0.9 - P_{chance}} \tag{A-6}$$

It is observed that some subjects do approach 1.0 probability with good imagery, but averages over a group of subjects do not. The subjects are trained and tested before the experiment, and the subjects are given rest periods. Prizes are awarded for the best performance, and this appears to motivate the subjects. Whether performance would improve or degrade in actual combat is not known. Certainly motivation would increase. However, these are difficult experiments, and it would seem that getting nine out of ten calls correct would indicate reasonable motivation on the part of the subjects. Whatever the source of these errors, they do not correlate to image quality.

As seen in Figure A-6, the TTP metric provides an excellent fit to the data. The new metric predicts accurately for various shape and size blurs, good and poor intrinsic target contrast, and various levels of noise. The average error is 0.046 and the maximum error is 0.21. The square of the Pearson's correlation (PSQ) is 0.94. Also, the sampling cutoff applied in the noise experiments does not affect model accuracy. Experiments 13, 19, and 20 had a half-sample frequency of 0.42 cycles per milliradian. The image content beyond the half-sample frequency was mainly aliased content and represented image corruption. To generate Figure 9, the integral for the TTP metric was taken from 0.0 to 0.42 cycles per milliradian. The TTP metric was not affected by a half-sample frequency cutoff.

The Johnson criteria are less accurate. In Figure A-5, there is a general scatter of the data. The PSQ is 0.72, the average error is 0.1, and the maximum error is 0.32. There is also a vertical line of values at N = 10.5 and again at N = 22 which result from limiting ξ_{cut} to the half-sample frequency. Figure A-7 shows results for the Johnson criteria without the half-sample limit. Prediction accuracy improves somewhat. For the experiments shown, the average error without the frequency limit is 0.096 and the maximum error is 0.36. The PSQ is 0.75.

Figure A-7. Johnson predictions without the half-sample frequency limit imposed.



A.3.2 Results of sampled imagery experiments

$N_{resolved}$ and $N_{JCresolved}$ are decreased by an amount that depends on the sampling artifacts predicted to be present in the image. The model used to predict the amount of sampling artifacts is described by Vollmerhausen (2000). Figures A-8 and A-9 show Experiment 25 results and model predictions for the Johnson criteria and the TTP metric, respectively. In both figures, the abscissa is $N_{resolved}$ (or $N_{JCresolved}$) and the ordinate is PID. As seen in Figure A-8, the Johnson criteria predictions are consistently pessimistic However, as seen in Figure A-9, the TTP metric does provide a good fit between model and data with a PSQ correlation above 0.9. Sampling predictions are pessimistic at long ranges (low metric values). This occurs because of the nature of the sampling correction. The correction is an empirically derived, fractional decrease in range performance. As the target gets further from the sensor and therefore smaller on the display, sampling actually has a greater impact on performance. However, this is currently not modeled.

Figures A-10 and A-11 show results from Experiment 36. Again, the Johnson criteria is pessimistic. The TTP metric accurately predicts performance with a PSQ of 0.93. In Figure A-11, TTP model predictions are accurate at long range but pessimistic at short range; this is the opposite of the Experiment 25 behavior shown in Figure A-9. Remember, however, that the subjects moved their heads, optimizing performance in a way not predicted by the model.

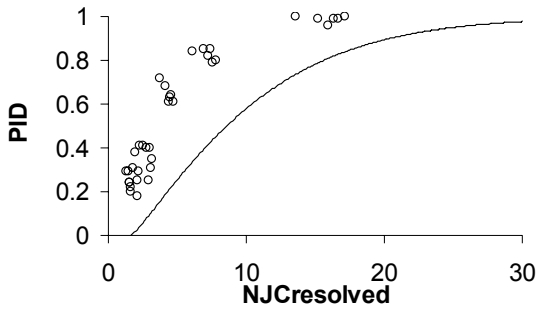


Figure A-8. Experiment 25 results and Johnson criteria model predictions.

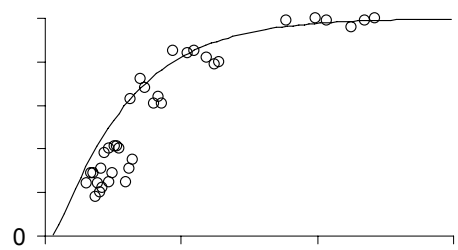


Figure A-9. Experiment 25 results and TTP model predictions.

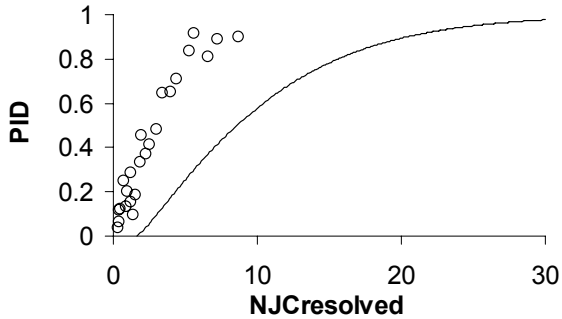


Figure A-10. Experiment 36 results and Johnson criteria model predictions

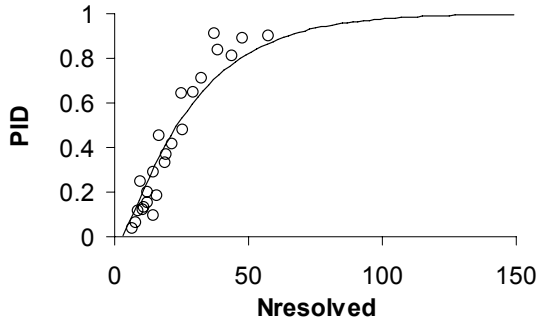


Figure A-11. Experiment 36 results and TTP model predictions.

A.3.3 Results of experiments with colored noise

These experiments were performed to illustrate that the TTP metric can be used to predict performance in the presence of colored noise. Figure A-12 shows the results for thermal imagery and Figure A-13 shows results for visible imagery. The N50 for the thermal images is 20.8, and the N50 for the visible images is 28. The TTP model fits the data well; the PSQ value is 0.9 in both cases.

Johnson criteria results are shown in Figures A-14 and A-15 for the thermal and visible images, respectively. The N50 for the visible images is 5 based on fitting the curve to the no noise and white noise data. There are systematic errors, particularly for the low frequency noise.

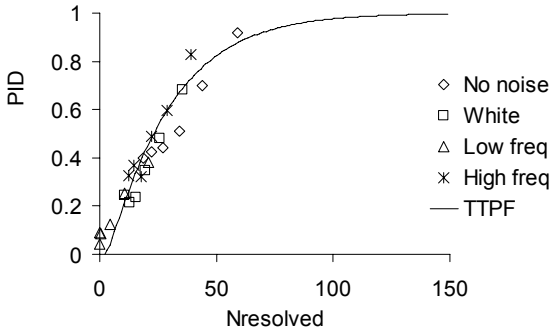


Figure A-12. TTP metric results of colored noise experiment with thermal imagery.

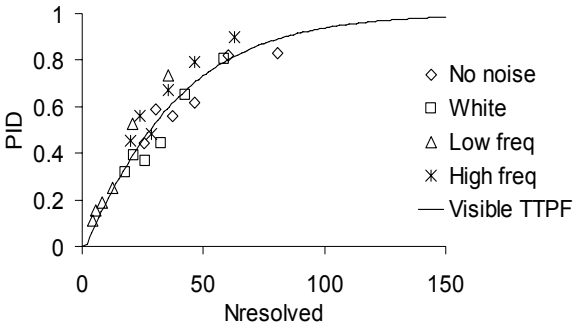


Figure A-13. TTP metric results of colored noise experiment with visible imagery.

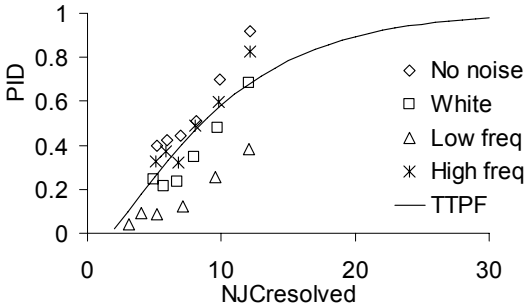


Figure A-14. Johnson criteria results of colored noise experiment with thermal imagery.

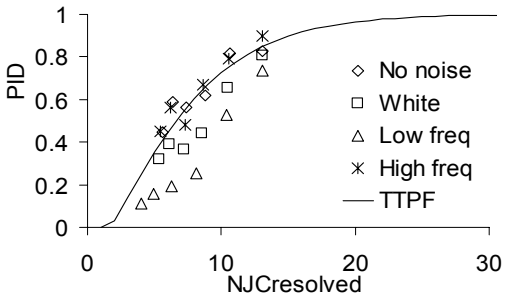


Figure A-15. Johnson criteria results of colored noise experiment with visible imagery.

Appendix B: Experiments with Low Contrast and Boost

In all of the validation experiments, the single largest error associated with TTP predictions occurred for low contrast (0.033), no-noise images with high-frequency boost applied. It initially appeared that the error might be systematic, so further evaluations were performed. The results of those evaluations provide some insights into the workings of the model and the pitfalls associated with this type of experimentation.

Experiment 34 used Gaussian blurs with $e^{-\pi}$ MTF cutoffs in object space of 0.2, 0.23, 0.27, 0.34, 0.46, and 0.69 cycles per milliradian. This was an ID experiment as described in Appendix A. The black and white display was used. The system magnification was 0.63. Since the images were minified compared to object space, frequency cutoff at the eye is proportionally greater than the cutoff in object space. The experiment consisted of applying the six Gaussian blurs to the 590 by 401 pixel, thermal images. Four sets of images were created, two with contrast of 0.11 and two with contrast of 0.033. One image set at each contrast had high frequency boost applied; see Figure B.1. No noise was added to the imagery. The imagery was down-sampled by four before presentation.

Figure B.1 Plot showing relationship between Gaussian blur, applied boost, and final "after boost" MTF.

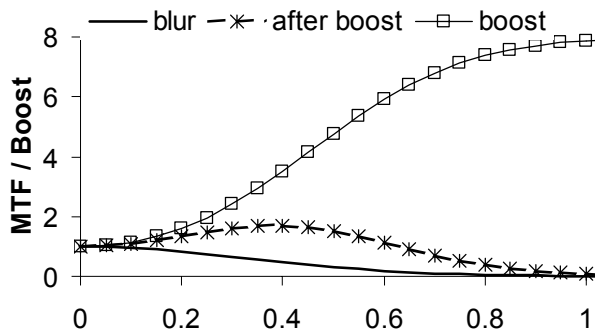


Figure B.2 shows the TTP predictions compared to the observer data. The data have been corrected for chance (0.083 probability) and for mistakes (0.1 probability). The largest error is for low blur, low contrast, with boost. There appears to be a systematic error for low contrast predictions, particularly with boost applied. The data is re-plotted in Figures B.3 and B.4. These figures show that the performance improvement due to boost is predicted well, but absolute predictions are pessimistic for low contrast when the blur is small.

Figure B.2 Results of Experiment 34 showing model and observer data.

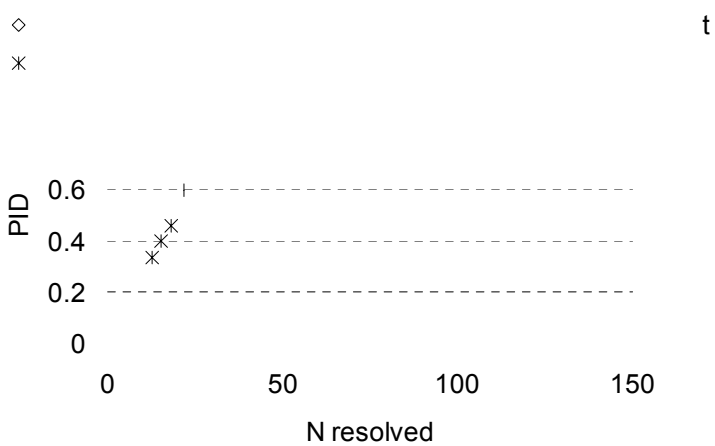


Figure B.3 Re-plot of the 0.11 contrast data and model. The "frequency cutoff number" refers to blur size, with small blur at cutoff 1.

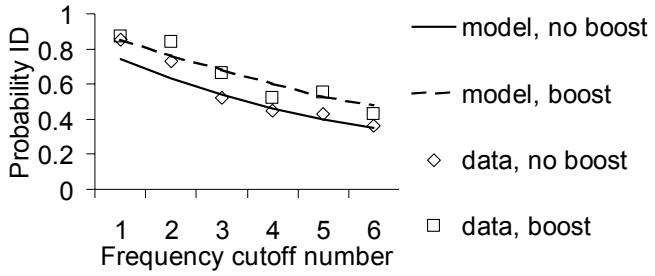
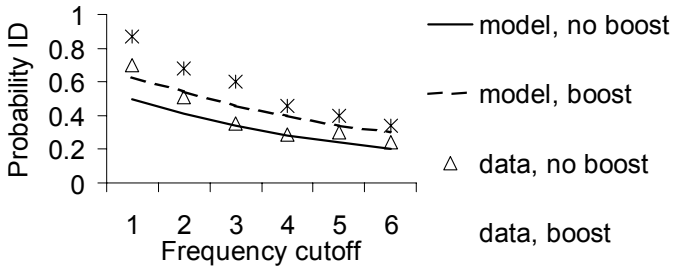
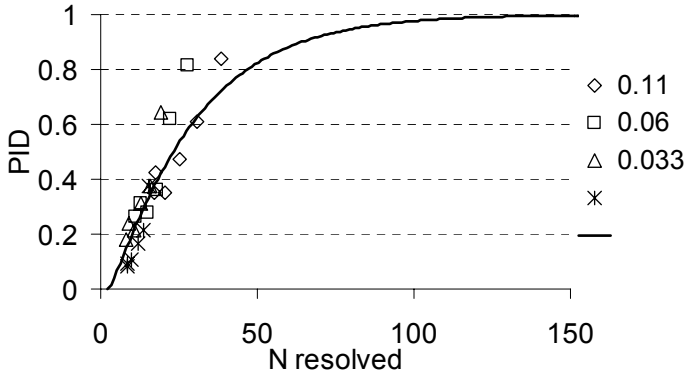


Figure B.4 Re-plot of the 0.033 contrast data and model. The "frequency cutoff number" refers to blur size, with small blur at cutoff 1.



Similar results were obtained for Experiment 33; details for this experiment are described in Appendix A. This experiment used Gaussian blur and explored the effect of changing contrast. Contrasts ranged from 0.11 to 0.018. The data are plotted in Figure B.5 Again, the model is somewhat pessimistic for the smaller blurs. In this case, boost was not used.

Figure B.5 Results of Experiment 33 showing model TTPF and observer data for 0.11, 0.06, 0.033, and 0.018 contrast target sets. The six data points for each contrast are different Gaussian blurs.



When the images from Experiments 33 and 34 were evaluated, it was determined that the cues needed to ID the targets were available in the small blur image sets. The data errors are not statistical. Remember that these images were not corrupted by noise, and that the display used had unusual contrast dynamic range (10 bit). The images associated with high observer probability did have good target cues.

To compare performance as contrast changes, the same target set is used as contrast as changed. In order to avoid many repetitions of showing the same image set, a different image set is used for each blur (each frequency cutoff). Experiment 39 was run to determine whether the model errors can be explained by a change in task difficulty. That is, if the model error is systematic, then changing the order in which blur is applied to the experiment cells should not affect the results. In Experiment 39, the cells which in Experiments 33 and 34 had small blur were given large blur.

The results of experiment 39 are shown in Figure B.6. The model pessimism did disappear. In Figure B.7, the results of Experiments 33 and 39 for the same contrast and blur are averaged. No attempt was made to establish the most difficult target set and average that with the easiest set; the matching occurred by chance. Clearly, a systematic model error does not exist.

Figure B.6 Results of Experiment 39 showing model TTPF and observer data for 0.11, 0.06, 0.033, and 0.018 contrast target sets. In this experiment, images which previously had large blur now had small blur.

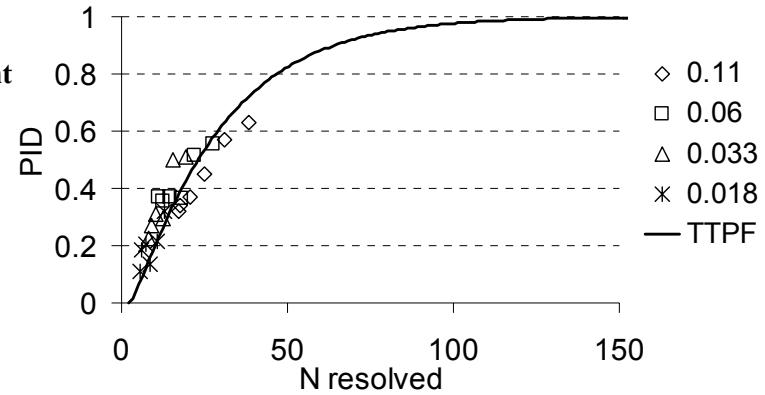
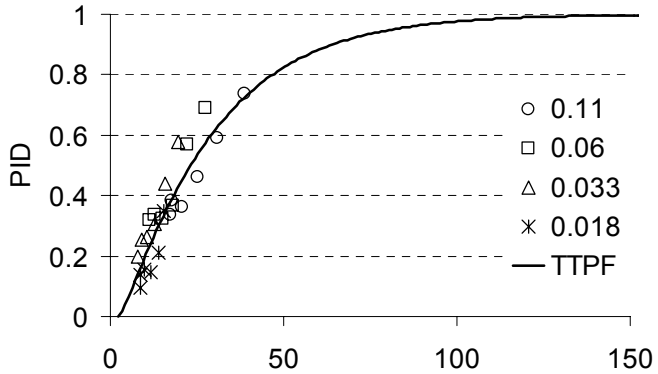


Figure B.7 Average of observer results from Experiments 33 and 39. Averaging task difficulty by mixing the targets viewed at each blur and contrast makes the model more accurate.



This evaluation points up two problems. First, target acquisition cannot really be *predicted* until we can predict task difficulty. The target is not in the model; we model image quality. Second, because human observers learn quickly, the same target image set cannot be used over and over. But comparing performance based on different target groupings leads to errors because of the change in task difficulty.

Appendix C: Recognition Experiment

The experiments used to develop the TTP metric used the ID task. The task was kept consistent so that the model did not change between experiments. With a fixed and known N50, the TTPF model curve is known and fixed. That same model curve is then compared to the results of numerous experiments, showing that the model can predict the impact of changing blur shape and size, noise, contrast, and sampling.

The primary reason for performing a recognition experiment is to verify that the sampling adjustments are applicable to an easier target acquisition task than ID. The recognition experiment is also a further check on the TTP metric.

Previously, target recognition involved discriminating between tanks, trucks, and armored personnel carriers (APC). An N50 of 3 for the Johnson criteria and 14.5 for the TTP metric is associated with this type of recognition task. However, trucks are much easier to discriminate from tanks or APC than APC from tanks. This recognition task is a mixture of easy and hard discriminations, and does not constitute a good target acquisition experiment for model validation.

Devitt (2001) describes a new recognition set consisting of tracked-armored, wheeled-armored, and wheeled-truck. She demonstrated that the three classes were equally difficult to discriminate. Further, this new recognition task has operational significance because wheeled combat vehicles are becoming more common. The target set used for this experiment is shown in Table C.1. Figure C.1 illustrates the three types of vehicles.

Figure C.1 Recognition Tracked-armored/Wheeled-armored/Wheeled-truck Experiment involved many vehicles and aspects; these are examples.

The Johnson N50 for the new recognition task is 3.5 (Devitt, 2001); the TTP N50 is therefore 16.9. The conversion between N50 values is discussed in Section 6. The square root of target area averaged over all targets and aspects is 2.93 meters. Average target contrast is 4.1 K.

A 256 by 256 focal plane array was used for this experiment. The detector pitch was 25 microns. The F/2 optics had a 7.33 centimeter focal length. Imagery was displayed on the black and white monitor. Simulated ranges were 0.43, 0.64, 0.97, 1.3, 1.6, and 2.15 kilometers. Various amounts and types of aliasing were created by changing detector active area (detector fill factor) and display technique. In-band aliasing was varied by changing the detector fill factor. Low in-band aliasing resulted from setting the detector active area to 25 microns (100% fill factor). High in-band aliasing resulted from setting the detective active area to 1 micron (fill factor of 1/25 in both directions).

To change out-of-band aliasing (visibility of pixels), different display interpolations were used in the experiment; these are shown in Table C.2. In all cases, sensor imagery was e-zoomed by 11 in both horizontal and vertical. Low out-of-band aliasing resulted from using the MATLAB bicubic image resize function to resize by eleven. The bicubic interpolation filtered out-of-band aliasing; no raster or pixel effects were visible. High out-of-band aliasing was created by using pixel replicate to e-zoom by eleven. In this case, the pixels were readily visible. The experiment consisted of four lines: (1) no in-band and no out-of-band aliasing, (2) no in-band with out-of-band, (3) in-band but no out-of-band, and (4) both in-band and out-of-band aliasing.

$N_{resolved}$ is decreased by an amount that depends on the sampling artifacts predicted to be present in the image; this is discussed in Section 7. The amount that $N_{resolved}$ is decreased for each line of the experiment is shown as the "sampling factor" in Table C.2.

Figure C.2 shows the observer data and the model predictions. The fit between model and data is excellent. Both the TTP metric and the adjustment for sampling artifacts are applicable to the recognition task.

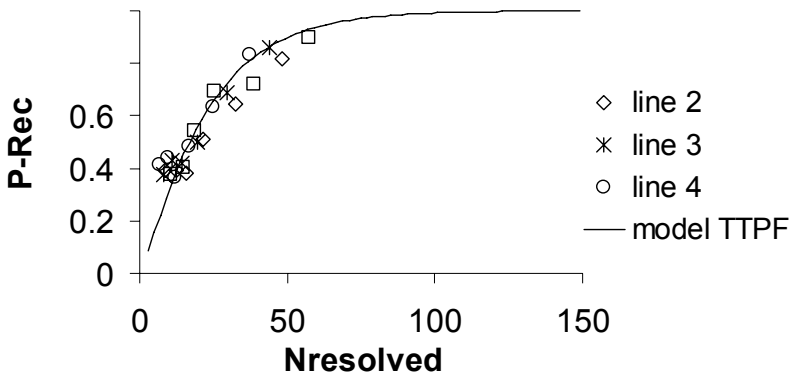


Figure C.2 Model (solid line) and observer data for recognition sampling experiment.

Table C.1 Vehicles and aspects used in the recognition experiment

TRACKED		WHEELED TRUCK (SOFT)		WHEELED ARMORED	
Target	Aspect	Target	Aspect	Target	Aspect
2S1	3NE, 7NG	HEMMT	0NG, 2NG, 4NG, 6NG	BRDM-2	4NE, 5NG, 1 DG, 6DG, 0NE, 2NG
2S3	0NG, 7NG,	M35	0NG, 2NG, 4NG, 6NG	BRDM-2 AT	0NE, 3NE, 5NE, 4NG, 6NE
ACRV	2NE, 6NG	ASTROS	3DG, 4NE, 5NE, 1DG, 6NG	BTR-70	4NE, 5NG, 1DG, 6DG, 1NG, 3NG, 7NG
AVLB	1DE	FMTV/Lt	3NE, 5NE, 7NE	LAV-25	0DG, 3DE, 6DG, 7NG
BMP-1	6NG, 7NE	FMTV/Md	0NG, 2NG	LAV-AD	0DG , 5DE, 1NG, 6NE
M1064	2NG	FROG-7	1DE, 4DG, 5DG, 1DG, 6DE	LAV-AT	0DG, 1DE, 2DE, 7DG
M109A5	1DG	GAZ-66	0NG, 3NE, 2NG, 5NE, 7NE	LAV-CC	4NE, 5NG, 2DG, 6DG
M109A6	3DE, 5NG	GRAD-1	4DG, 5DG, 1DG, 6NE, 7NE	LAV-M	4DG, 3DG, 2NG, 7DE
M113	4NG	HMMWV	0NG, 3NG, 2NE, 7NG, 1NE	LAV-Rc	4DE, 5NG, 7DG, 3DE
M1A1	0NE	HMMWV -TOW	6NE, 7NE, 1DG, 4DE, 5NG	M-93A1	0NG, 3DE, 2NE, 7DG, 2DE, 5NE
M1IP	4NG, 5DE	STYX	4NG, 5NE, 1DE, 6DG, 3NE		
M2	2NE, 3NE				
M48	1DG				
M548	2NE, 3DG				
M551	4NE, 1NE				
M577	0NE, 7DG				
M578	5DE				
M60A3	0NG, 7NE				
M728	0DE, 6DG				
M88	5DG, 7DE				
M901	4DE, 5DE				
M992	4NE				
MTLB	3NG, 6NE				
T-55	2NG, 6NG				
T-62	4DG, 1NE				
T-72	3NE, 6NG				
T-72 (Reac)	1NG				
ZSU-23/4	0NG, 2NE				
M41	5DE				

ASPECT KEY

FIRST CHARACTER

0= front, 1=left front, 2=left flank, 3=left rear, 4=rear, 5=right rear 6=right flank, 7=right front

SECOND CHARACTER

N= night 8-12 micron thermal
D= day 8-12 micron thermal

THIRD CHARACTER

G = 0 elevation
E = 7 degree elevation

Table C.2 Interpolations, fill factors, and $N_{resolved}$ sampling factor for each experiment line

line	interpolate	E-zoom	System magnification	Detector fill-factor	Sampling factor
1	Bicubic	11	10.6	Large	0.96
2	Replicate	11	10.6	Large	0.8
3	Bicubic	11	10.6	Small	0.73
4	Replicate	11	10.6	Small	0.62

Appendix D. ID Performance with speckle imagery

The ability of the TTP metric to predict performance of humans viewing images produced by laser range gated (LRG) imagers was investigated. A perception experiment was designed and an image set was simulated. The simulated sensor was modeled on current electron bombarded CCD (EBCCD) technologies. Because the phenomenology of laser range gated sensors is a mix of coherent and incoherent effects, both types of processes had to be represented in the imaging chain. Figure D.1 shows where in a representative imager the transfer characteristics of the sensor are calculated as field intensity (coherent) or power (incoherent).

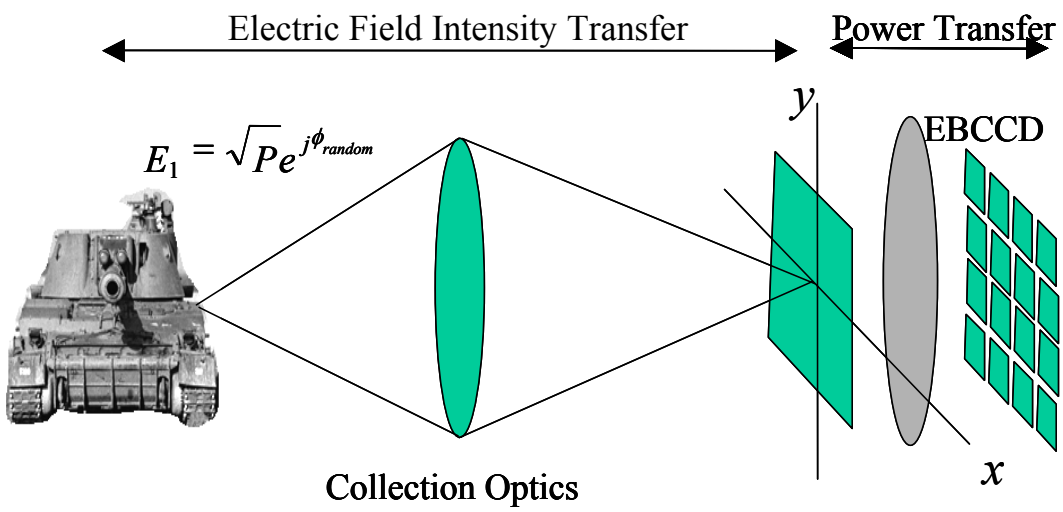


Figure D.1 Imaging chain in LRG sensor

To simulate the coherent portion of the imaging chain, the field from the object was propagated through the collection optics and imaged onto the image plane of the camera. Taking the square root of the gray level in a panchromatic image of the target simulated the amplitude of the coherent field incident on the sensor aperture. The phase of each pixel in the object was chosen from a uniform distributed random variable over the interval $[0,2\pi)$ since the target was considered to be rough compared to the wavelength of the laser illumination. For each point on the source object, a coherent impulse response (blur) was created in the image plane. Since the source points had random phases, the resulting image was formed through the interference of a number of impulse responses at the image plane. The resulting output field at the image plane was the complex input (electric field) convolved with the coherent impulse response. The field was then converted to irradiance by squaring the magnitude of the field at the focal plane. All other blurs in the sensor, including the electron proximity focus and detectors of the EBCCD, were linear with respect to the irradiance and were applied as point spread functions in power.

The characteristics of the simulated sensor are given in Table D.1. Using these characteristics, the coherent and incoherent impulse response functions for the optics and detectors were created.

Table D.4. Simulated sensor parameters.

Parameter	Value
Wavelength	1.57 microns
Aperture diameter (centrally obscured)	125 mm
Aperture obscuration fraction	.3125
Focal length	1250 mm
Detector size	13 microns square

The coherent impulse response or point spread function (PSF) of the optics was calculated using the following equation

$$PSF_{Optics}(r) = somb\left(\frac{r}{\Omega_0}\right) - \varepsilon^2 somb\left(\varepsilon\frac{r}{\Omega_0}\right) \tag{D.1}$$

where r is an angular subtense measured from the sensor and Ω_0 is the ratio of the aperture diameter to the focal length. The incoherent impulse response of the electron proximity focus was implemented as a filter in the spatial frequency domain. The filter function was modeled by fitting a supergaussian to measured MTF data. The resulting MTF is given by

$$MTF_{Prox}(f) = \exp\left\{-\left|\frac{f}{\beta}\right|^{\gamma}\right\} \tag{D.2}$$

where γ was found to be 1.64 and β was found to be 25.5 cycles per milliradian of angle measured from the sensor. The incoherent detector PSF was calculated using

$$PSF_{Det}(r) = rect\left(\frac{r}{w}\right) \tag{D.3}$$

where w is the angular subtense of a detector measured at the sensor. After application of all blurs, the imagery was downsampled by a factor of two, which resulted in imagery having 295 horizontal pixels and 200 vertical pixels.

The system described above was simulated under four conditions. The first condition was incoherent (spatial and temporal incoherence). Under this condition, no random phase was applied to the pristine image. The second condition was a single shot LRG-SWIR mode where temporal coherence was maintained and spatial phase was randomized, thus creating a speckle image. The third condition was a two-pulsed average image and the third condition was an eight-pulsed average image. The averaging decreased the effect of the laser speckle.

The sensor simulation was applied to 576 images that were presented to U.S. Army soldiers as part of a perception experiment. The primary purpose of the experiment was to determine the impact of speckle on target identification performance. The standard NVESD identification target set is shown in Figure D.2. "Probability of identification" (PID) was established by NVESD as the ability of an observer to identify one of these targets from the other eleven targets. The target set included these 12 targets with 12 aspects resulting in 144 pristine images that were processed four different ways to produce the 576 perception test images. The targets were chosen for their relative confusability and tactical significance. The left flank of each vehicle in the visible target set is shown in Figure D.2. The visible images were collected using 35mm cameras with color film at a range of 25 meters. The film was digitized, converted to grayscale, and processed to have a resolution of 1.8 cm per sample in both horizontal and vertical directions. The images were used as the source power that was converted to an electric field in the simulation. Examples of the simulated images are shown in Figure D.3.

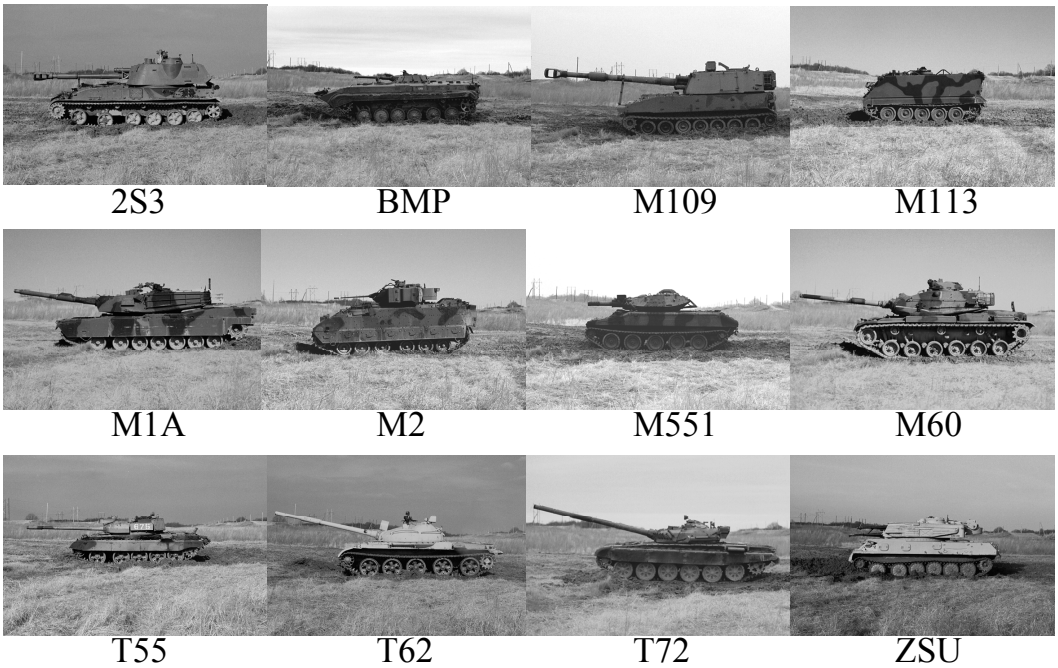


Figure D.2. Target set used in perception experiment.

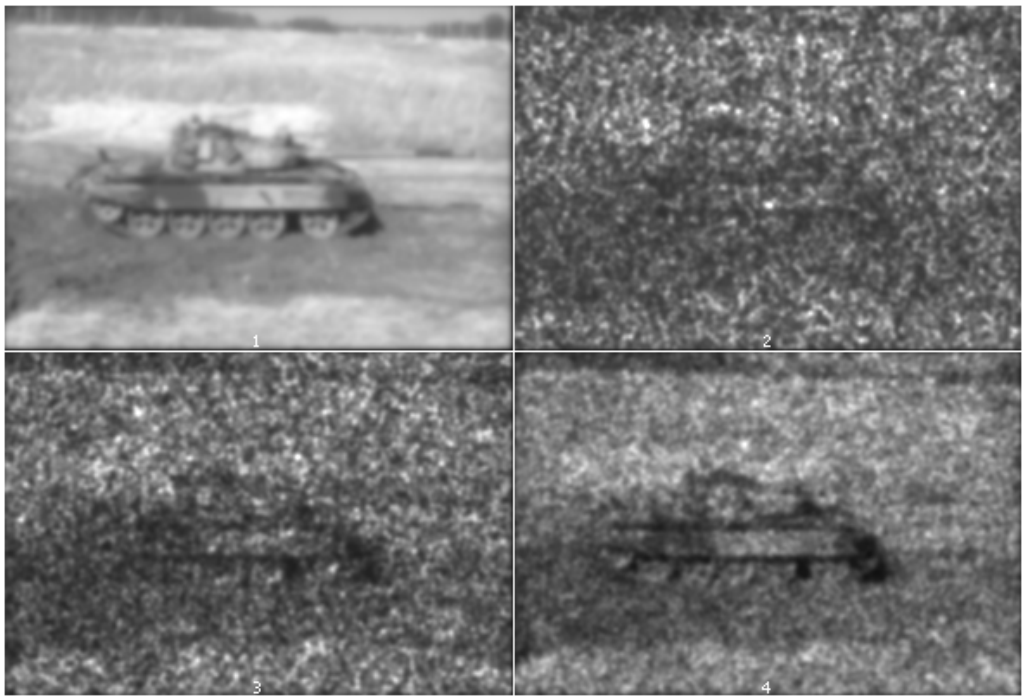


Figure D.3. Simulated speckle images at 5km. 1. Incoherent. 2. Coherent – no averaging. 3. Coherent – 2 speckle average. 4. Coherent – 8 speckle average.

The perception experimental design is outlined in Table D.2. From the pristine image sets of Figure D.2, 12 targets with 12 aspects (144 images total) were distributed evenly across the columns of the table shown. This distribution resulted in only two of the same target image in each column and two of the same aspect in each column yielding 24 pristine images associated with each column. The 24 images were processed with a prescribed range. In the column labeled "1," 24 pristine images were blurred to correspond to a 1-km range through an incoherent sensor model and placed in cell "AA." Also, 24 pristine visible images (the same targets and aspects as those in AA) were processed with a 1-km range through a single laser range-gate shot sensor model and placed in cell "AB" and so on.

Table D.2. Experiment Design

	Range(km) 1	3	5	7.5	10	15
Incoherent	AA	BA	CA	DA	EA	FA
Single Shot	AB	BB	CB	DB	EB	FB
Two Shot Average	AC	BC	CC	DC	EC	FC
Eight Shot Average	AD	BD	CD	DD	ED	FD

The standard NVESD target set was processed in a manner depicted in Table D.2. A comparison of speckle images is shown in Figure D.3. The images shown were simulated at 5 km with an incoherent process, a single-shot laser pulse, a 2-shot image average, and an 8-shot image average. Image averages were obtained by adding independent speckle images. The incoherent image was taken as a baseline for imagery comparison. It can be easily seen that the speckle from the single-shot image can degrade the identification performance significantly.

Fifteen soldiers were trained to identify targets with 99% proficiency prior to participating in the experiment. The experimental cells were randomized to vary the level of target identification difficulty. The images were displayed on high-resolution grayscale monitors. The monitors have Gaussian MTFs with equivalent spot sizes of 0.306 mm horizontally and 0.237 mm vertically. There were approximately 70.9 pixels per centimeter on the displays. The images were displayed with an average luminance of 5 fL and were viewed from a nominal distance of 15 inches. After correcting for chance guesses and mistakes, the average probability of ID for each cell in Table D.2 is given in Table D.3. The probability of a chance guess was 1 in 12 and the probability of mistake was 1 in 10.

Table D.3. Percent Correct Identification

	A	B	C	D	E	F
A	96.6%	99.5%	81.2%	83.3%	87.7%	72.4%
B	99.0%	67.7%	26.2%	15.8%	10.9%	1.6%
C	99.7%	71.0%	40.1%	21.3%	17.9%	5.1%
D	99.3%	83.7%	64.7%	45.9%	32.7%	20.2%

In order to predict the performance for each cell in the perception experiment, the contrast threshold function (CTF) of the system (including the observer) had to be computed. The system CTF would then be used to compute the target task performance (TTP) metric from which the probability of identification (PID) could be derived. To simplify these performance predictions, it was assumed that all the spatial distortions of the sensor could be applied incoherently and that speckle acted as additional display noise.

In computing the system CTF, distinctions must be made in the definition of spatial frequency. This is due to the different ways of measuring angles in the system. From a sensor point of view, angles can be determined by the size of the target and the range. From the observer's point of view, angles are determined by the size of the target on the display and the observation distance. The two angles are related by the magnification of the system. The relationship between the two spatial frequencies is given by

$$f_t = M * f_e \tag{D.4}$$

where f_t is spatial frequency measured at the sensor, M is the magnification, and f_e is spatial frequency measured at the eye. In this experiment, all images were displayed at the same size with the blur from the system increasing with increasing range. Therefore, for the image sizes and display characteristics used in this experiment, the relation between range and magnification was given by

$$M = 0.01028 * R \tag{D.5}$$

where R is the range in meters.

The system CTF was computed in spatial frequency measured at the eye. The equation used for the system CTF is given by

$$CTF_{SYS}(f_e) = \frac{CTF_{eye}(f_e)}{MTF_{SYS}(M * f_e)} \sqrt{1 + \alpha^2 \frac{\sigma^2}{L^2}} \tag{D.6}$$

where CTF_{eye} is the CTF of the observer's unaided eye, MTF_{sys} is the system MTF, α is a calibration constant which has been found to be 169.6, σ is the standard deviation of the display noise, and L is the average display luminance. The system MTF is found from

$$MTF_{SYS}(M * f_e) = MTF_{Optics}(M * f_e) MTF_{Prox}(M * f_e) MTF_{Det}(M * f_e) MTF_{Disp}(f_e) \tag{D.7}$$

The MTF of the proximity focus (MTF_{Prox}) is given by equation D.2. The incoherent MTF of the optics is given by

$$MTF_{Optics}(f) = \frac{A + B + C}{1 - \varepsilon^2} \tag{D.8a}$$

where

$$A = \begin{cases} \frac{2}{\pi}\left[\cos^{-1}\left(\frac{f}{f_0}\right) - \frac{f}{f_0}\sqrt{1 - \left(\frac{f}{f_0}\right)^2}\right] & f \le f_0 \\ 0 & f > f_0 \end{cases} \tag{D.8b}$$

$$B = \begin{cases} \dfrac{2\varepsilon^2}{\pi}\left[\cos^{-1}\left(\dfrac{f}{\varepsilon f_0}\right) - \dfrac{f}{\varepsilon f_0}\sqrt{1-\left(\dfrac{f}{\varepsilon f_0}\right)^2}\right] & f \le \varepsilon f_0 \\ 0 & f > \varepsilon f_0 \end{cases} \tag{D.8c}$$

$$C = \begin{cases} -2\varepsilon^2 & f < \dfrac{1-\varepsilon}{2} f_0 \\ \dfrac{2}{\pi}\left[\varepsilon \sin\phi + \dfrac{\phi}{2}\left(1+\varepsilon^2\right) - \left(1-\varepsilon^2\right)\tan^{-1}\left[\dfrac{1+\varepsilon}{1-\varepsilon}\tan\left(\dfrac{\phi}{2}\right)\right]\right] & \dfrac{1-\varepsilon}{2} f_0 \le f \le \dfrac{1+\varepsilon}{2} f_0 \\ 0 & f > \dfrac{1+\varepsilon}{2} f \end{cases} \tag{D.8d}$$

The parameters ϕ and f_0 are given by

$$\phi = \cos^{-1}\left[\frac{1+\varepsilon^2 - \left(\dfrac{2f}{f_0}\right)^2}{2\varepsilon}\right] \tag{D.8e}$$

and

$$f_0 = d/\lambda \tag{D.8f}$$

where d is the aperture diameter. The detector MTF is given by

$$MTF_{Det}(f) = \mathrm{sinc}\left(f/w\right) \tag{D.9}$$

where w is the angular subtense of the detector. The display MTF is assumed Gaussian and given by

$$MTF_{Disp}(f) = \exp\left[-(\rho f)^2\right] \tag{D.10}$$

where, for f measured in eye space, ρ is the angular subtense of the Gaussian blur size of the display measured at the eye.

The display noise is found from

$$\sigma^2 = \frac{\sigma_h^2 \sigma_v^2}{N} \tag{D.11}$$

where N is the number of frames averaged (1,2, or 8),

$$\sigma_h^2 = \int_{-\infty}^{\infty} S_n(v) \left| MTF_{Det}(v) MTF_{Prox}(v) MTF_{Dizp}(v) MTF_{Eye}(v) MTF_{Per}(v) \right|^2 dv \,, \qquad \text{(D.12)}$$

and

$$\sigma_v^2 = \int_{-\infty}^{\infty} S_n(v) \left| MTF_{Det}(v) MTF_{Prox}(v) MTF_{Dizp}(v) MTF_{Eye}(v) \right|^2 dv \,. \qquad \text{(D.13)}$$

$S_n(v)$ is the power spectral density (PSD) of the noise. For speckle, the PSD is found from

$$S_n(v) = L \frac{MTF_{Optics}(v)}{\int_0^{\infty} MTF_{Optics}(u) du} \qquad \text{(D.14)}$$

where L is the average display luminance.

Two new MTFs are introduced in Equation (D.12) and (D.13). The first is the MTF of the eye which is given by

$$MTF_{eye}(v) = MTF_{eye_optics}(v) MTF_{retina}(v) MTF_{tremor}(v) \qquad \textbf{(D.15)}$$

where

$$MTF_{eye_optics}(v) = \exp\left[-\left(\frac{av}{M} \right)^b \right] \qquad \textbf{(D.16)}$$

$$MTF_{retina}(v) = \exp\left[-0.375 \left(\frac{v}{M} \right)^{1.21} \right] \qquad \textbf{(D.17)}$$

$$MTF_{tremor}(v) = \exp\left[-0.4441 \left(\frac{v}{M} \right)^{2} \right] \qquad \textbf{(D.18)}$$

The values of a and b in equation (D.16) are found from

$$a = 43.69 \left[\exp\left\{ 3.663 - 0.0216 D_{pupil}^2 \ln\left(D_{pupil} \right) \right\} \right]^{-1} \qquad \text{(D.19)}$$

and

$$b = \left(0.7155 + \frac{0.277}{\sqrt{D_{pupil}}} \right)^2 \qquad \text{(D.20)}$$

with D_{pupil} being the diameter of the pupil. Pupil diameter is dependent upon light level and can be calculated from

$$D_{pupil} = -9.011 + 12.23 \exp\left\{ -\frac{\log(L)}{21.082} \right\} \quad (mm) \qquad \text{(D.21)}$$

where L is the luminance in foot-Lamberts.

The second new MTF in equation D.16 is a perceptual filter that describes the spatial frequency response of the visual cortex channel and is given by

$$MTF_{Per}(v) = \exp\left\{-2.2\log\left(\frac{v}{f}\right)\right\} \tag{D.22}$$

where f is the frequency at which the CTF is being evaluated. This filter is only applied in the horizontal direction.

Once the system CTF is calculated for a given range, the TTP metric can be computed. The TTP metric is given by

$$TTP = \frac{\sqrt{A_t}}{R}\int_{f1}^{f2}\sqrt{\frac{C_t}{CTF_{SYS}(f_e)}}df_e \tag{D.23}$$

where $f1$ and $f2$ are defined as the points where the value C_t intersects CTF_{SYS}. The average size of a target in pixels was 7021. With an observation distance of 38.1 cm and approximately 70.9 pixels per cm on the display, the target critical dimension (the leading term in equation (D.23)) was 31 mrad. Using this value, equation (D.23) was evaluated for every cell in the test matrix. The results are shown in Table D.4.

Table D.4. TTP Metric Values for Perception Experiment.

	A	B	C	D	E	F
A	193.29	143.12	106.16	76.97	58.50	37.44
B	75.28	31.97	18.79	10.98	6.76	2.09
C	89.86	23.84	23.84	14.93	10.06	4.71
D	121.87	59.19	36.88	23.98	17.16	10.03

The metric values in Table D.4 were converted to PID using the target transform probability function (TTPF). The TTPF is given by

$$P = \frac{\left(\frac{TTP}{V_{50}}\right)^E}{1+\left(\frac{TTP}{V_{50}}\right)^E} \tag{D.24}$$

$$E = 1.33 + 0.23 * \left(\frac{TTP}{V_{50}}\right)$$

where V_{50} is the value of the metric necessary for the task to be performed 50% of the time. For identification, the value of V_{50} has been determined to be 28.

Using the values from table D.4 with the TTPF, the PID values were predicted. These values are compared with the measured PID values in Figure D.4

The TTPF explains approximately 94% of the variance in the measured data.

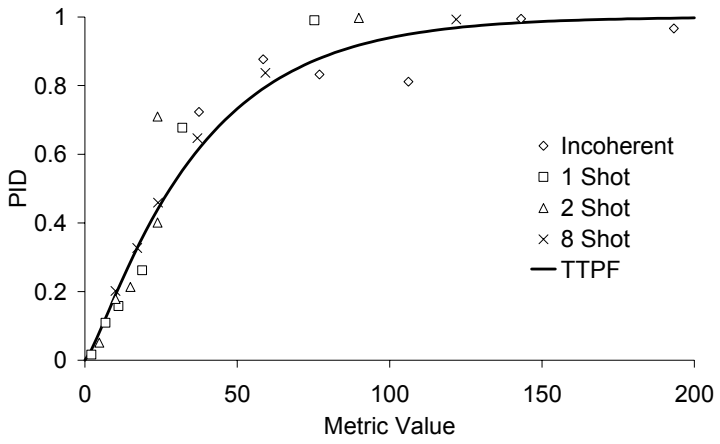


Figure D.4. Comparison of predicted and measured PID values.

Appendix E. Model Details

1. EYE MTF

The eye MTF is taken from Overington (1976). The eye MTF includes factors for refraction optics, retina, and tremor; each of these is calculated with a numerical fit to the data in Overington. Eye MTF varies with light level because it depends on pupil diameter. See Table E–1 for pupil diameter versus display light level. These pupil diameters are for one eye; the pupil diameter for two eyes is about 0.5 mm smaller.

Table E-1. Pupil Diameter Versus Light Level

Pupil Diameter (mm)	7.0	6.2	5.6	4.9	4.2	3.6	3.0	2.5
Light Level (LOG fL)	-4	-3	-2	-1	0	+1	+2	+3

fr = frequency at the eye in cycles per milliradian

bb = LOG10(display luminance)

dpul = -9.011 + 13.23 * EXP(-bb / 21.082)

dpul = dpul - (eye# - 1) * .5

e0 = (.7155 + .277 / dpul ^ .5) ^ 2

fi = EXP(3.663 - .04974 * dpul ^ 2 * LOG10(dpul))

fe = 43.69 * fr

eye MTF = EXP(-(fe / fi) ^ e0) * EXP(-.375 * (fr) ^ 1.21)* EXP(-.4441 * fr * fr)

2. CONTRAST THRESHOLD FUNCTION

The numerical approximation provided by Barten (1992, 2000) is used to predict sine-wave CTF.

$$\mathrm{CTF}(u) = (a\, u\, EXP(-b\, u)\, (1 + c\, EXP(b\, u))^{0.5})^{-1}$$

$$\mathrm{num} = 540\, (1 + 0.7/L)^{-0.2}$$

$$\mathrm{denom} = 1 + 12/(w\, (1 + u/3)^2)$$

$$a = \mathrm{num}/\mathrm{denom}$$

$$b = 0.3\, (1 + 100/L)^{0.15}$$

$$c = 0.06$$

where

u = Spatial frequency in cycles per degree,
w = Square root of picture area in degrees, and
L = Display luminance in cd/m^2.

The CTF values for a monocular display are increased by the √2 [see Volume I, Section 1.802 in Boff and Lincoln (1988)].

3. CHARGE-COUPLED DEVICE MTF

The following parameters are used in the equations:

D_{pit} = Pitch of detector in milliradians,

D_{fil} = Fill factor,

V_{int} = Spacing of pseudo interlace in milliradians ,

C_{TE} = Charge transfer efficiency,

N_H = Number of charge transfers,

E_{cut} = Electronic cutoff as a fraction of $1/(2 \cdot H_{pitch})$,

Ep = Number of poles in electronic filter, and

F_{req} = Spatial frequency in cycles/mrad.

Dimensions in milliradians are found by taking a thousand times the dimension in centimeters, multiplying by the fiber-optic taper reduction ratio, and then dividing by the objective focal length.

The MTF associated with spatial integration by the detector is

$$MTF_{detector} = sin(\pi D_{pit} D_{fil} F_{req}) / (\pi D_{pit} D_{fil} F_{req}).$$

This applies both horizontally and vertically. The horizontal MTF associated with the electronic sample and hold is

$$MTF_{SAMP/HOLD} = sin(\pi D_{pit} F_{req}) / (\pi D_{pit} F_{req}).$$

In addition to the electronic sample and hold, a CCD camera normally has an electronic filter. The filter roll-off is applied to the horizontal and is modeled as:

$$MTF_{filter} = 1 / (1 + (F_{req} / (E_{cut} / (2 H_{pitch})))^{2EP})^{1/2}.$$

A CCD normally has both horizontal and vertical transfer registers. The MTF loss due to a charge transfer efficiency less than 1 is modeled as follows.

$$MTF_{CTE} = N_H (1 - C_{TE}) (1 - cos(2 \pi D_{pit} F_{req})).$$

Pseudo interlace involves adding adjacent vertical detectors starting with detector row 1 on the first field and with detector row 2 for the second field. During video field 1, the first video line consists of the signals from detector row 1 and detector row 2 added together. The second video line in field 1 consists of detector row 3 and detector row 4 added together, etc. During video field 2, the signal on the first video line consists of detector row 2 and detector row 3 added together. The vertical MTF associated with pseudo interlace is

$$MTF_{interlace} = cos(\pi V_{int} F_{req}).$$

4. MTF OF FIBER-OPTIC TAPER

According to Schott (Siegmund, 1989), the limiting resolution of a fiber-optic taper is approximately 600 divided by the fiber pitch in μm. Assume that limiting resolution occurs at 3-percent contrast, and that the MTF is Gaussian. If taper input end pitch is in μm and objective focal length is in centimeters, then:

$$MTF_{reducer} = EXP\,(-C_{oef}\,F_{req}^{2})$$

$$C_{oef} = 10{,}000\ Pitch^{2}\ LOG_{e}\,(.03)\ /\ (600\ F_{0})^{2}.$$

5. OUTPUT BRIGHTNESS FOR I^2 CCD AND CRT DISPLAY

The output brightness for an I^2 CCD and display is

$$B_{disp} = D_{gain}\,(B_{out}\,C_{volt}\,G_{CCD} + K)^{gama} + L_{min}$$

where:

B_{out} is defined above except EP_{tr} is reducer transmission,
C_{volt} = volt out of CCD per footcandle input,
G_{CCD} = Gain of CCD AGC,
D_{gain} = fL out of display per input (volt)^{gama},
$gama$ = gamma (intensity power law exponent), and
$K = (L_{min}/D_{gain})^{1/gama}$.

6. DIFFRACTION-LIMITED OPTICS MTF

The diffraction-limited optics MTF is given by:

$$MTF_{dif} = (2/\pi)\ (\cos^{-1}(Q) - (Q)\ (1 - Q^{2})^{1/2})\ or$$

$$MTF_{dif} = (2/\pi)\ (\tan^{-1}(Q') - (Q)\ (1 - Q^{2})^{1/2})$$

where

$Q = \lambda \cdot F_{req}/D$,
λ = Wavelength in μm,
F_{req} = Spatial frequency in milliradians,
$D = F_0/F\#$ (Optics aperature diameter in millimeters), and
$Q' = (1 - Q^2)^{1/2}/Q$.