

Resolution requirements for alphanumeric readability

John Lester Miller, MEMBER SPIE
Research Triangle Institute
4000 Kruse Way Place
Lake Oswego, Oregon 97035
E-mail: Jlmiller@RTI.org

John M. Wiltse, MEMBER SPIE
FLIR Systems, Incorporated
16505 Southwest 72nd Street
Portland, Oregon 97224

Abstract. Multiple series of data were acquired specifically to relate the number of resolvable cycles across an observation of an alphanumeric character to enable humans to accurately read it. This has serious implications for the resolution needed for a surveillance camera to present a “readable” image to a human. We present the theory based both on an extension of the Ratches/Johnson criteria and Fourier analysis. The theory is supported by empirical data based on user identification of random English letters and Arabic numerals. The results strongly indicate that accurate readability (defined as 90% correctness or better) can be accomplished with approximately 2.8 cycles across a block letter. This appears to suggest a lower requirement than that generally accepted for unknown target identification. However, this is consistent due to the limited dataset: there are only 36 alphanumeric characters, and the observer intuitively possesses this *a priori* knowledge. Moreover, the ability to read an alphanumeric is a steep function of the resolution at a certain number of cycles, which lies between 2 and 3 cycles. The authors define a new level of discrimination, “read,” which lies between the classical “recognize” and “identify.” The probability of correct reading can be expressed similarly to that of Detection, Recognition, and Identification by using a postscript such as “Read₉₀.” © 2003 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.1542892]

Subject terms: alphanumerics; recognition; identification; resolution; electro-optical imaging systems; modeling.

Paper 020231 received Jun. 5, 2002; revised manuscript received Aug. 23, 2002; accepted for publication Aug. 27, 2002.

1 Introduction

1.1 Theory

There is increasing desire for surveillance systems to provide image content and quality that allows the user to be able to read alphanumerics for positive registration identification. We use the term “read,” rather than the stricter traditional terms of “recognize” or “identify,” because of the specific meanings attached in the community to those terms. By read, we mean to be able to properly identify a random English, high-contrast block letter or traditional Arabic numeral with high accuracy. The set of possible characters includes only the 26 English alphabet characters and the ten Arabic numerals in common use in English-speaking countries, for a total set of only 36 possible unique symbols. English-speaking people use these symbols every day and perform this function many times per day, changing their perceived resolution by changing their distance (or range) while attempting to read something. Thus, because this set is not infinite, and people are highly trained at identifying a letter with minimal cycles, intuitively it should be possible to accurately identify an alphanumeric character with fewer cycles than it would be to identify an object from a larger, less familiar set.

To resolve alphanumeric characters (or any other target, for that matter), the imaging system should not significantly low-pass filter the spectral contents of the image. Knowing

the spectral contents of the alphanumeric characters gives an initial estimate of the resolution requirements for the sensor. Five randomly chosen letters were selected: B, A, R, Y, and W. The font was uppercase, Times (a serif font), nonbold. We used two-dimensional Fourier analysis to compute the spatial frequency content of each of these characters. The results are shown in Fig. 1. Note that there is little spectral power above 3 cycles/letter height (LH) in the vertical direction. Indeed, with the exception of the letter B, there is very little spectral power at vertical spatial frequencies above 2.5 cycles/LH. It is interesting to note, however, that there is much more high-frequency spectral power in the horizontal direction than in the vertical direction. This analysis suggests that being able to resolve anything more than 3 cycles per letter height is unnecessary to read the characters. Even if there is no *a priori* knowledge, Fourier analysis indicates that alphanumerics should be readable with about 3 or 4 cycles of resolution. This may have implications for machine vision.

1.2 Background

There have been numerous efforts and publications since the initial work of Ratches, Johnson, Rosell, Gerhart and others to quantify the minimum number of cycles required across a military target for identification.¹⁻⁶ Based on two-

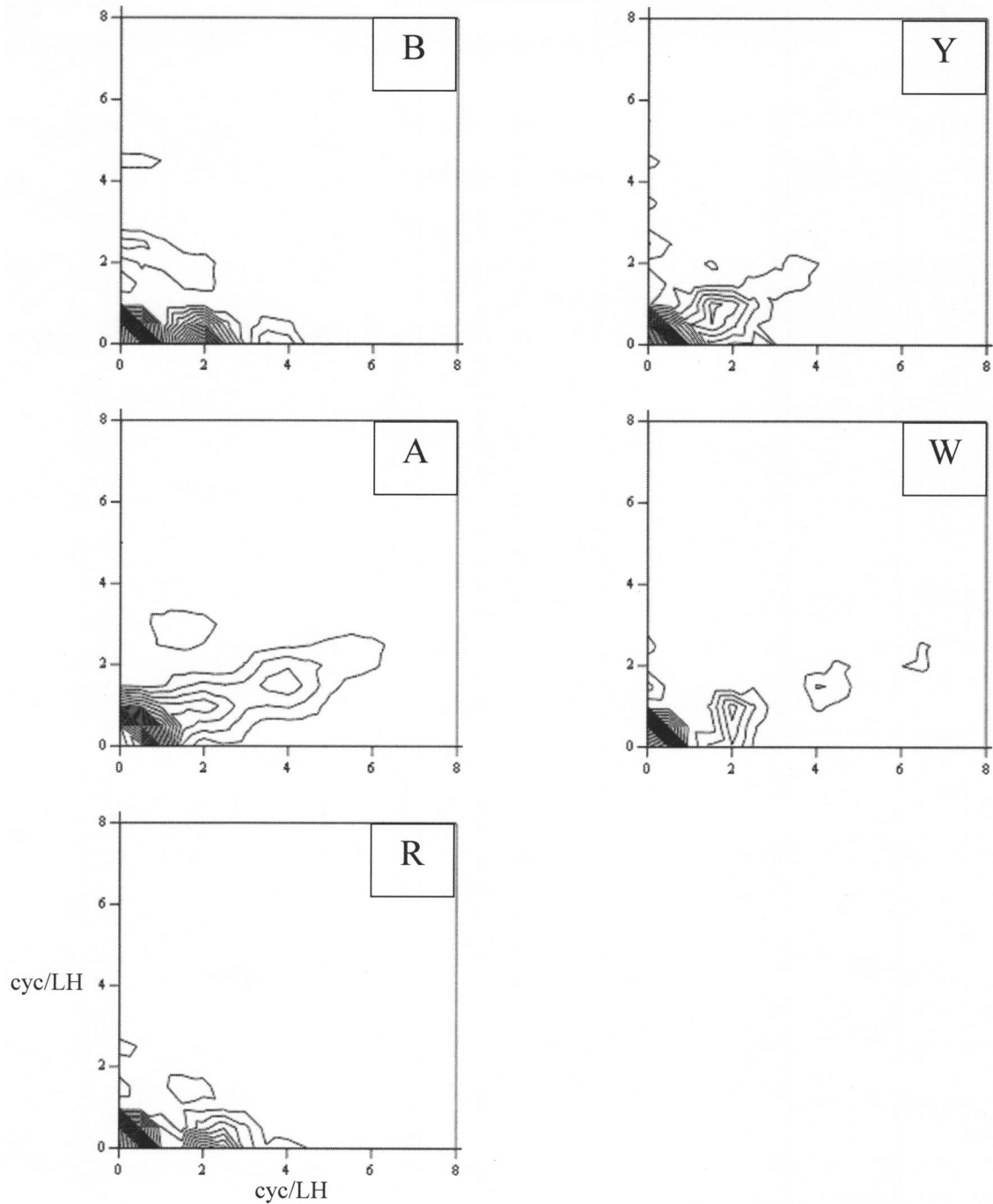


Fig. 1 Two-dimensional Fourier transforms of the randomly chosen letters *B*, *A*, *R*, *Y*, and *W*.

dimensional sampling, a high contrast displayed resolution required for detection, recognition, and identification is given in Table 1.

The Johnson discrimination methodology is the standard process by which an object is assigned to a subset of a larger set of objects based on the detail perceived by the observer. In this methodology, the number of cycles across the target critical dimension is used to divide visual discrimination into criteria levels: detection (knowing that an object is present), recognition (knowing the class to which

Table 1 Generally accepted criteria for detection, recognition, and identification.

Function	Number of cycles across one dimension	Number of cycles across critical dimension for two dimensions
Detection	1	0.5 to 0.75
Recognition	4	3 to 4
Identification	8	6 to 8

an object belongs), and identification (knowing the specific type within that class).

Johnson originally developed his criteria to discriminate among various types of military targets: soldiers, tanks, airplanes, etc. Observers discriminated between the various targets, and the equivalent bar pattern frequency was determined. Johnson's original testing was driven by image intensifiers, but since that time his methodology has been extensively used for all types of imagers as well. Johnson's experiments and those that followed mainly focused on discriminating large, military targets—those that would be seen in combat operations, often by infrared sensors. This same methodology was applied to the ability of a user to read alphanumeric.

Specifically when considering alphanumerics, Smith⁷ claims that the following resolution limits are required to photograph printed or typewritten copy: eight line pairs per the height of a lowercase letter “e” for excellent reproduction (including serifs, etc.), five line pairs per letter height for easily legible reproduction, and three line pairs per letter height for decipherable reproduction.

1.3 Motivation

Reading and deciphering alphanumeric text is important in many surveillance and law enforcement applications. In many cases, the legibility of a ship, airplane, or automobile registration makes the difference between evidence and no evidence. Since these registration markings are always significantly smaller than the vehicle they identify (e.g., the cross section of a car may be 5×1.5 m, whereas each number on its license plate may only be 3×7.5 cm), this can become the limiting factor in evidence gathering. While many sensors can detect, recognize, and identify the entire vehicle, typically only sensors operating in the visible or near-IR can see the reflectivity differences between types of paint that makes alphanumeric text visible.

As high-end IR and electro-optical (EO) sensors move from purely military to paramilitary (customs service, border patrol) and police applications (airborne, and even ground-based for traffic control), they are required to do more than simply detect, recognize, and identify targets. They are required to provide evidence for law enforcement, and positive identification for military targeting and tracking. Moreover, urban conflicts as recently experienced by Israel, NATO forces in Bosnia, and U.S. antiterrorist forces often require locating and tracking vehicles by registration. In this new paradigm, infrared or wide field-of-view EO sensors are used to detect, recognize, and identify targets, while narrow field-of-view EO sensors operating in the visible or near-IR are required to gather alphanumeric data for evidentiary and pursuit purposes. These sensors can be standard CCD television cameras, intensified CCDs, low-light CCDs, or active television systems.^{8,10}

Compared to the vast amount of literature on classification of military targets, there is little data available on the resolution requirements for alphanumeric readability. Providing data on the resolution requirements is crucial for systems analysis and design of narrow field-of-view EO systems used to gather this information. Since systems engineers must often trade the competing requirements of narrow field-of-view (which provides high spatial resolution) with a wider field-of-view (which provides more contextual

information, as well as an easier control for pointable and/or stabilized systems), it is important to know how many cycles must be resolved across alphanumeric critical dimensions to ensure readability.

The purpose of this work is to determine the resolution requirements needed to discriminate between different alphanumeric characters. This should be considered the limits of decipherability: at the “read” limit, characters are difficult to see and it is not possible to resolve serifs, discriminate between different types of fonts, etc.

1.4 Assumptions

We are concerned with the 26-letter English alphabet and the 10-number Arabic numeral set. Note that the English alphabet (as well as several others) is derived from the Greek alphabet, which contains 24 characters. Results for any of these character sets, assuming the observer is fluent in that set, should be similar.

In contrast, the resolution requirement for classifying the thousands of discrete symbols in some eastern languages is likely different than that for classifying Greek, English, or other western European alphabets.

The data in our tests are based on a human's ability to interpret an image captured by an electro-optical sensor. There is no contrast enhancement employed or automatic character recognition software, although one would expect that these could greatly improve results. Merely, the image is displayed or printed for a human to comfortably view and attempt to interpret based on their life experience. In all cases, the humans know that they are viewing an alphanumeric. Common nonletter language symbols such as @, &, +, and punctuation were not part of this study.

It is always assumed that users are allowed to change the image magnification by altering the distance from their head to the display or paper to optimize perception based on their eyes and the displayed media.

Some alphanumeric characters are easier to distinguish from others. For example, it is harder to distinguish between an F and an E or an O and a Q than it is a 1 and a W. Specific pairings such as these were not considered, and are expected to be part of a future study. Except in some rare legal cases, the confusion of one letter from a specific other letter is not thought to be of serious concern, as the registration of a vessel consists of multiple characters. If one can read six out of eight characters, there is a good chance that the correct registration can be determined.

The data for this research consisted of random sequences of alphanumerics, as the impetus was on being able to read registration, which frequently consists of such. In all cases, the targets were random characters, not complete words. One could assume that reading common words or numbers would require even fewer cycles, as one only needs to read a majority of the letters in a word to read the word. For example, reading a known word should be easier as it requires fewer correctly identified letters to identify a word (e.g., the reader of this probably knows what we mean if he or she sees “Alph num ric,” although two letters are missing).

This is different from the task of automatically classifying well-resolved characters, for example, in handwriting recognition or optical character recognition (OCR). In those cases, resolution is not the issue. The issue is automating

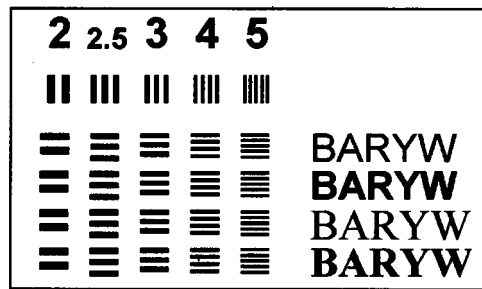


Fig. 2 Alphanumeric target board.

the human character recognition process so it can be performed effectively by a computer.

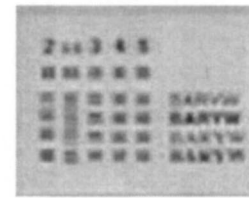
2 Methodology, Field Tests, and Results

2.1 Over-Sampled Images

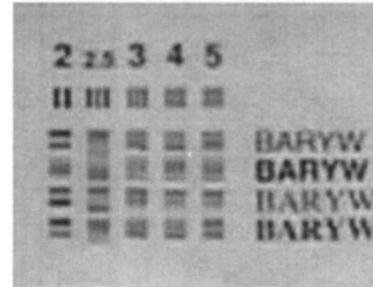
A simple laboratory test was conducted to determine the resolution requirements for alphanumeric readability. An alphanumeric target was created and printed with a laser printer (Fig. 2). This target had the randomly chosen letters B, A, R, Y, and W. The letters were repeated with the following font combinations (top to bottom): Helvetica (a sans serif font) regular and bold, and Times (a serif font) regular and bold. Beside each character group are five bar patterns corresponding to 2, 2.5, 3, 4, and 5 cycles per character height. Note that we used vertical resolution (corresponding to a character's height), and not horizontal resolution (corresponding to its width). All the characters in this image are of the same height, while their width varies significantly depending on the letter (i.e., A is much narrower than W).

The target was mounted on a wall of the laboratory. Approximately 6 m away, a commercial zoom CCD block camera (a Sony FCB-IX470) was used to videotape the target board. The camera's autofocus feature was turned on. The camera's analog Y-C video was recorded using a Sony MiniDV digital tape recorder. Still frames were acquired with an analog frame grabber. Note that the particulars of the camera, VCR, and frame grabber selection are not particularly significant. This is because the bars are printed on the same target board as the letters, so any MTF loss is apparent in both the bars and the characters. The zoom position of the camera was adjusted until each of the five groups of horizontal bars could be barely resolved (the vertical bars were not used). Video was recorded and one frame from each zoom position was captured.

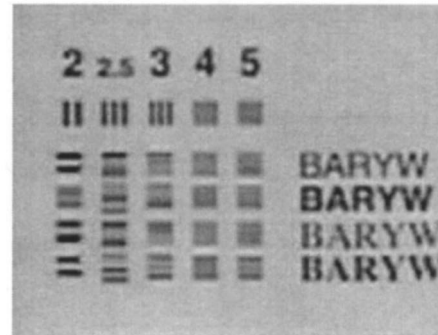
The results of this test are shown in Fig. 3. Figure 3(a) shows the case in which 2 cycles per letter height (LH) could be resolved on the laboratory monitors (note that the resolution is further degraded by the frame capture and printing process). Clearly, the text is illegible. Figure 3(d) shows the letters with 4 cycles/LH resolution. At this resolution, the letters are very easily legible with no ambiguity. Finally, Figs. 3(b) and 3(c) show the target at approximately 2.5 cycles/LH and 3 cycles/LH, respectively. At 2.5 cycles/LH [Fig. 3(b)], the letters are mostly legible. Some errors are still possible—for example, the B might be taken for an 8, particularly with the regular Times font (third from top). However, by 3 cycles/LH [Fig. 3(c)], all the



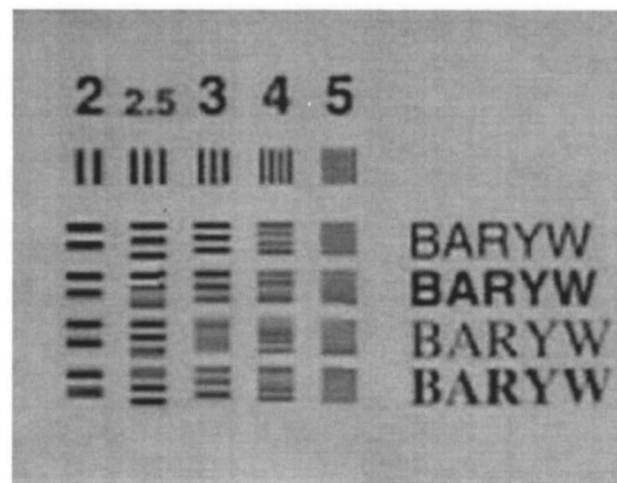
(a)



(b)



(c)



(d)

Fig. 3 Images of the target board at various approximate resolutions: (a) 2 cycles/letter height (LH), (b) 2.5 cycles/LH, (c) 3 cycles/LH, and (d) 4 cycles/LH.

letters in all font combinations are clearly legible. This test was repeated outdoors in field trails with similar results.

2.2 Pixilated Images

In cooperation with the Air Force Research Laboratories at Kirtland Air Force Base, New Mexico, we conducted tests

of reflecting monochrome 810 nm light off target boards. The target boards consisted of white block letters painted on a battleship-gray background. The letters were 15.4, 30.8, and 45.7 cm in height. As mentioned before the height of the characters was held constant but the width varied slightly from one character to another. A sunlit example is shown in Fig. 4.

We digitally acquired data using a Graflex variable focal length lens assembly, PixelVision Camera, and computer frame grabber at various ranges using several focal lengths. This provided us with a multitude of pixels across the height. The letters were illuminated with monochrome laser light of varying brightness. All of the data presented here were with highly illuminated targets, and the images are all of high brightness and easily readable from a signal-to-noise perspective, so the brightness is not a variable here. It has been well known that contrast and noise can affect the ability to discern a target and these results will be different for low-contrast images.

A high-contrast printed version of these images was generated (Fig. 5 is an example), and 50 people (who were not part of the experiment) were asked to identify a series of alphanumeric characters with various number of pixels across the height. The printed test sheets resulted in various pixels across the height of the target, randomly oriented as to where the pixel began. This is a key point: no attempt was made to align the camera for optimal pixel phasing, and thus the Kell factor applies (see next). The human subjects consisted of a random sample across FLIR Systems, Kirtland Air Force Base, and others. All had English as their first language. None of these people were given any *a priori* knowledge of the alphanumerics (other than it was an English alphabet letter or numeral) and none were involved in the field tests or knew what the target boards looked like. The respondents were instructed to only guess at one letter. A few guessed that it might be one of two alphanumerics (e.g., they indicated that it was either a C or a O); these were considered incorrect readings. When a respondent left a blank (failed to guess at the alphanumeric), it was also considered an incorrect reading.

The results of this survey are presented in Fig. 6. These results are for pixilated, undersampled images. The three curves are for the three unique alphanumerics the subjects were required to identify, knowing only that they were either a letter or a numeral. Generally, all the alphanumerics that we requested the subjects to read fell into similar curves, although the authors concede that there would be certain combinations of letters that required more or less resolution to identify.

Consistent with previous studies of military targets, the curve is generally an S curve. It is interesting to note the steepness of the curve between two and three cycles. This suggests that there is either a psychophysical human response, or an inherent structure of English letters, contributing to the results (or both). At a critical resolution (around three cycles), identifying alphanumerics becomes easy, while one cycle of less resolution makes this task very difficult and inaccurate. It is not a linear function. Obviously, it is easy to discriminate between diverse characters such as Q and I with few cycles across them, representing the lower part of the S curve. Likewise it is much more difficult to

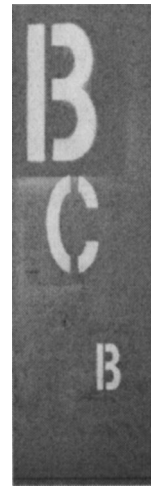


Fig. 4 Example of alphanumerics use for pixilated tests.

discriminate between Q and O, and requires a much greater resolution, representing the top plateau of the S curve.

The data of Fig. 6 were corrected to yield cycles across target by multiplying the spatial frequency by 0.7, the well-documented Kell⁹ factor accounting for losses for converting pixilated still images to oversampled video images. The Kell factor is used because although two samples theoretically can perfectly sample a sinusoid, the output is heavily dependent on phase. In the worst case the sinusoid is completely missed, whereas with three samples, the sinusoid is well sampled regardless of the sample phase (see Fig. 7).

3 Conclusions

Our results indicate that 2.5 cycles/LH is sufficient for decipherability, and that at a resolution of 3.0 cycles/LH the text is easily readable.

Both the over-sampled and under-sampled results are surprisingly consistent, as they both indicate that it is pos-

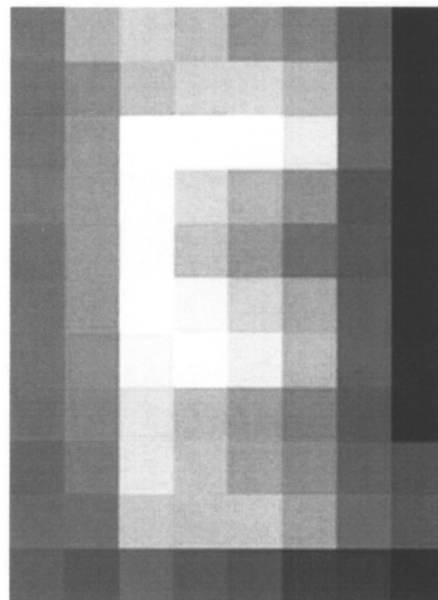


Fig. 5 Example of a "pixilated E." This example has eight pixels across the height or four cycles.

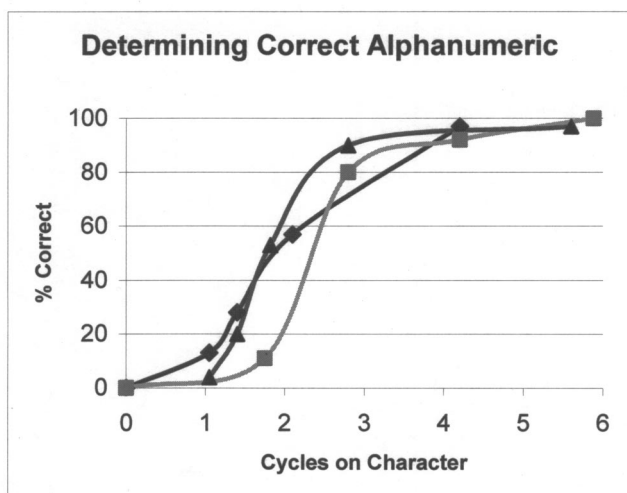


Fig. 6 The number of correct readings of an alphanumeric as a function of cycles for the letters C (marked by diamonds), F (marked by triangles), and B (marked by squares).

sible to read English alphanumerics with high accuracy with approximately 2.5 to 3.0 cycles across the height of the alphanumeric.

Additional resolution makes the reading more pleasant and potentially less fatiguing, but not significantly more accurate. Conversely, less resolution makes the reading almost impossible. Slight increases or decreases in resolution outside of the critical resolution zone (two to three cycles) matter little to final results.

When considering alphanumerics, the authors would suggest that an additional criterion be established and used, which lies between the classic “recognize” and “identify,” and is termed “read₉₀.” It is implied that this criteria signifies a 90% positive knowledge of an alphanumeric character in the native language of a literate observer. The 90% threshold puts it above the critical resolution for most letter combinations (able to correctly identify a C from an O) and requires about three cycles or six pixels on a character.

Acknowledgments

The authors would like to thank the USAF Air Force Research Laboratories at Kirtland Air Force Base for supporting this work, as well as FLIR Systems, The Research Triangle Institute, The New Mexico State Police, The German Navy, and Optimare Incorporated. We’d also like to acknowledge the personal support and hard work of Jim Byars, Chuck Crompton, Bruce Ellison, Dan Gwynne, Rob Ireland, John Kelly, Dick Kerr, Dagmar Kinne, Dennis McCall, Scott Way, and Stewart Worley.

References

1. J. Ratches, “Static performance model for thermal imaging systems,” *Opt. Eng.* **15**(6), 525–530 (1976).
2. M. Cantella, “Staring-sensor systems,” in *Infrared and Electro-Optical Systems Handbook*, Vol. 5, S. Campana, Ed., pp. 149–189 (1993).
3. K. Seyrafi, *Introduction to Electro-Optical Imaging And Tracking Systems*, pp. 183–192, Boston: Artech House, Boston, MA (1993).
4. J. Miller and E. Friedman, *Photonics Rules of Thumb*, pp. 48–49, McGraw Hill, New York (1996).

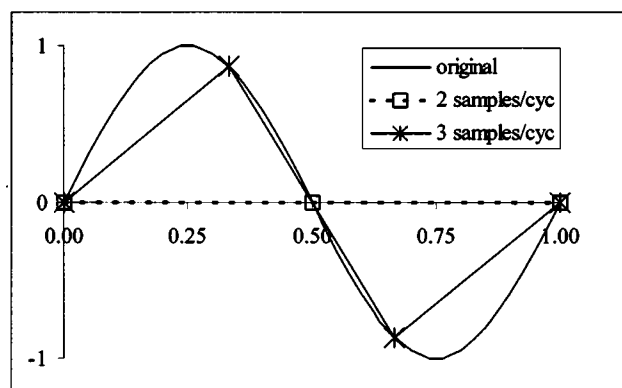


Fig. 7 A sinusoid sampled with 2 samples/cycle and 3 samples/cycle. Although the former meets the Nyquist criterion, results depend strongly on the phase of the sampling with respect to the signal.

5. J. Miller, *Fundamentals of Infrared Technology*, pp. 408–410, Kewler Publishing, New York (1993).
6. J. Johnson, “Analysis of image forming systems,” *Proc. Image Intensifier Symp.*, pp. 249–273, Warfare Electrical Engineering Dept., US Army Engineering Research and Development Laboratories, Ft. Belvoir, VA (1958).
7. W. A. Smith, *Modern Optical Engineering*, p. 355, McGraw-Hill, New York (1990).
8. J. L. Miller, J. M. Kelly, R. J. Ireland, and J. E. Byars, “Performance of a gimbaled fiber optic coupled laser spotlight and active television,” *Opt. Eng.* **37**(10), 2703–2709 (1998).
9. P. G. Barten, *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality*, pp. 181–182, SPIE Press, Bellingham, WA (1999).
10. J. L. Miller, J. M. Kelley, and J. Ehlen, “Flight test and performance of a nongated active television system,” *Proc. SPIE* **3698**, 250–254 (1999).



John Lester Miller received his BSc in physics from the University of Southern California and MBA from Regis University, and has completed graduate courses at the University of Hawai’i and Cal State Long Beach. In the past 22 years, he has held positions as research and development manager, program manager, functional manager, lead engineer, and electro-optical engineer on multispectral sensors, space surveillance sensors, astronomical instruments, image processing algorithm development, diamond films, missile warning systems, helmet-mounted cameras, and active television systems. He has written more than 25 papers and two books (*Principles of Infrared Technology*, Kluwer, ISBN 0-442-01210-1 and *Photonics Rules of Thumb*, McGraw-Hill, ISBN 0-070044329-7) on infrared technology (and two more to be published in 2003). He recently served as chief scientist at FLIR Portland and currently manages the Portland, Oregon office of the Research Triangle Institute.



John M. Wiltse received his BS in engineering and applied science from the California Institute of Technology in 1988. He received his MS in aerospace engineering in 1989 and his PhD in mechanical engineering in 1993, both from the University of Arizona. He has written four papers on active flow control. He developed inkjet printers for three years at Tektronix. For the last four years, he has worked on electro-optical and infrared imaging systems at FLIR Systems, where he is a Senior Systems Engineer.