LAW ENFORCEMENT STANDARDS PROGRAM

IMAGE QUALITY CRITERION FOR THE IDENTIFICATION OF FACES

prepared for the
National Institute of Law Enforcement and Criminal Justice
Law Enforcement Assistance Administration
U.S. Department of Justice

by

Gary T. Yonemura
Center for Building Technology
National Bureau of Standards

MAY 1974

U.S. DEPARTMENT OF JUSTICE
Law Enforcement Assistance Administration
National Institute of Law Enforcement and Criminal Justice

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 - Price 65 cents
Stock Number 2700-00261

ACKNOWLEDGMENTS

This document was prepared by the Law Enforcement Standards Laboratory of the National Bureau of Standards under the direction of Marshall A. Isler, Program Manager for Security Systems, and Jacob J. Diamond, Chief of LESL.

CONTENTS

FOREWORD

Following a Congressional mandate* to develop new and improved techniques, systems, and equipment to strengthen law enforcement and criminal justice, the National Institute of Law Enforcement and Criminal Justice (NILECJ) has established the Law Enforcement Standards Laboratory (LESL) at the National Bureau of Standards. LESL's function is to conduct research that will assist law enforcement and criminal justice agencies in the selection and procurement of quality equipment.

In response to priorities established by NILECJ, LESL is (1) subjecting existing equipment to laboratory testing and evaluation, and (2) conducting research leading to the development of several series of documents, including national voluntary equipment standards, user quidelines, state-of-the-art surveys and other reports.

This document, LESP-RPT-0303.00, Image Quality Criterion for Identification of Faces, is a report on an experiment which was designed and conducted by LESL for the purpose of determining the relationship between the human ability to identify faces in an image and the performance of imaging equipment such as night vision devices. The experiment was part of the LESL effort to develop standards and guidelines for night vision devices. Additional reports, as well as other documents will be issued under the LESL program in the areas of protective equipment, communications equipment, security systems, weapons, emergency equipment, investigative aids, vehicles and clothing.

Technical comments and recommended revisions are invited from all interested parties. Suggestions should be addressed to the Program Manager for Standards, National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U. S. Department of Justice, Washington, D. C. 20530.

Lester D. Shubin, Manager
Standards Program

*402(b) of the Omnibus Crime Control and Safe Streets Act of 1968, as amended.

IMAGE QUALITY CRITERION FOR IDENTIFICATION OF FACES

1. EXECUTIVE SUMMARY

The performance required of imaging devices should be based on the needs of the user. In analyzing these needs, three psychophysical levels of visual task performance must be considered. The simplest level is detection, where the observer is required to determine the presence of an object: a disk, line, gap, etc. The second is recognition, where the observer's task is not only to detect the presence of an object, but also to recognize it as belonging to a group or class; for example, airplanes, humans, dogs, etc. The third and highest level is identification, which includes the preceding two tasks, but has the additional requirement of discrimination between items within a given group: for example, identifying an individual person, a specific type of aircraft, etc. A paradigm of the above discussion is: I see something (detection), the object is a human (recognition) and his name is John Doe (identification). The experimental data base for these levels of performance is very uneven: many detection studies have been performed, fewer have addressed problems of recognition and rarer still are investigations associated with identification. The purpose of this study was to determine experimentally observers' perception of the image quality required for the identification of faces, as determined by two criteria: (1) the average observer, and (2) 90 percent of the population. The subjective response from the human observer was then transformed into a physical descriptor amenable to direct measurement by instruments.

Front and profile photographs of 40 individuals (white males) were taken under identical conditions; a knife edge and three-bar crenelate type gratings at 25 spatial frequencies were also photographed under the same conditions. The photographs of the people were shown to a panel of observers who were asked to choose the ten faces which most closely resembled one another. In the preliminary phase, one of the ten faces was degraded by defocusing into 16 levels of subjectively equal degradation steps. A small group of observers (12) was used to determine the approximate degradation levels appropriate for identification purposes and to make judgments to define differences between degradation levels which were approximately equal. The results of this preliminary experiment were used to degrade the photos used in the final experiment.

These ten faces, each degraded at 12 subjectively equal steps, were shown to 144 observers who picked the level of degradation that was judged as "good enough" for identification purposes. Each observer was asked to evaluate all ten sets of photos. The average acceptable degradation level and the level accepted by 90 percent of the sampled population were determined. The corresponding degraded knife edges were microphotometered and quantified into modulation transfer function and acutance measures. Based on relevant experimental studies available in the literature and the findings of this study, these two indices are recommended as alternative criteria for specifying the required performance of imaging devices.

The minimum acceptable performance for different face sizes and viewing distances is presented in tabular and graphical form. The procedures and background information necessary to apply these techniques are included.

2. INTRODUCTION

In the assessment of man's performance with imaging devices, two disciplines are most directly involved: physics (stimulus) and psychology (response). The physics of imaging devices has been extensively investigated and engineering design of these devices is proceeding satisfactorily. But unless improvements intended for the user do in fact lead to user benefits, the effort is nonproductive. For example, it is generally agreed by researchers in the field of image quality that the minimum resolvable bar target (which is the test most often used in lens quality studies) does not correlate with subjective evaluation of image quality [1–4]. The physical descriptor of the performance of an imaging device must be based on the response of the human eye.

In extracting information from the visual field, the psychophysical performance required of the observer depends on the type of visual information required. These visual tasks differ significantly in difficulty and complexity. At the simplest level is detection—determining the presence or absence of any object in the field of view. The next higher order of visual task performance requires the observer not only to detect the presence of an object, but also to recognize it as a member of a group. This level will generally require shape or form discrimination [5]. "The object I see is a triangle, a man, an animal, a gun, etc." The level of visual performance requirement is called recognition. Detection and recognition as followed by identification. This level requires prior performance of the first and second levels, and in addition the observer must be able to discriminate between members of the same group. The additional performance requirement is the perception of details; for example, features, type of gun. etc. A simple paradigm of the above discussion is: I see something (detection), it is a man (recognition), and his name is John Doe (identification).

2.1 *Psychological Factors*

All three levels of visual task performance discussed earlier are influenced by past sensory experience, criterion levels, attention, attitude, stress, motivation, etc. [6]. These psychological factors, become increasingly important as the difficulty of the required visual performance is increased from detection to identification, and are assumed to be included in the obtained judgments of image quality. No attempt will be made in this paper to discuss the other psychological factors in detail, except for those factors directly relevant to the conduct of this study. For example, in preliminary studies the observers frequently commented that their response would be dependent upon the importance of their reply. This and the other psychological variables were handled by allowing the subjects to react as they saw fit. The subjects were instructed to "Choose the photograph you feel is just good enough to identify the individual shown on the top leaf," (which contained an undegraded photograph). When the observers asked the experimenter to define "just good enough," they were told that one of the experimental objectives was to determine how individuals defined "just good enough to identify."

An important parameter is the confidence of the response. During the preliminary phase of this study, a photo of John F. Kennedy was used to determine degradation techniques and approximate image degradation levels. The amount of defocusing required before the observers stated they could not recognize the photograph as John Kennedy was surprising. We attribute this ability to identify grossly degraded photos as being based on past experience; that is, familiarity with the face of a famous individual. A study by Harmon [7] is relevant to the above observation. He asked his subjects to match names with highly blurred portraits. Since only names were given, the observers had to be familiar with the individuals portrayed and had to rely on past sensory experience to match each face with the appropriate name. The observers were able to identify highly blurred faces which at short distances looked like cubist paintings. Of the 14 faces used in the study, one was identified 96 percent of the time, the least identifiable was correctly identified only 10 percent of the time, the average being 48 percent correct identification. Harmon's findings on the relation between confidence of rating and accuracy of response are especially interesting and support the image quality criterion used in this study. He found that when the confidence rating was low the ratio of incorrect to correct responses was about 9:2, but at the highest confidence rating the ratio of correct to incorrect responses was 11:1. The correlation between an observer's confidence rating and his error response indicates that the observer's subjective evaluation of image quality is related to identifiability. In summary, a highly degraded image can be identified if the individual to be identified is well known to the person making the judgment. The converse is also true; a degraded picture will be erroneously identified as a popular or familiar individual. In addition, the error rate drops significantly when the confidence rating is high. This implies that what the observer judges as being a good picture will correlate well with minimum incorrect identifications.

It is true that when an observer is forced to make a choice or guess, he will be able to identify photos significantly more degraded than those he would subjectively choose as being adequate, but the error rate will be high. Furthermore, it is the observer's opinion that counts. Even if by a forced-choice technique we prove that a given percentage of the people can identify faces at highly degraded levels, it would still be difficult to convince some people that they should "guess" because they would probably be right. In this study the purpose was to obtain statistical data on the image quality level chosen as being acceptable for identification purposes. To this end the criterion for image quality was a photo that was judged to be just good enough to identify the individual pictured in a high quality photograph.

What are the psychophysical correlates of identification? There is general agreement that image quality cannot be completely described by a single physical parameter [8]. Part of this difficulty lies in the individual definition of image quality. A case in point is the study by Stultz and Zweig [2]. They used the method of paired comparison (observers are asked to compare photographs, two at a time, and state which has more of the desired quality), and found that when the observers were asked to judge the pictures on the basis of "picture quality," sharpness and granularity were given approximately equal weights. But when these same observers were asked to judge the pictures for "definition," the results correlated highly with sharpness and gave a low correlation with granularity. This is analogous to the intentional defocusing often used by portrait photographers to obtain an esthetically pleasing portrait as opposed to a sharply focused image beneficial for identification.

2.2 *Psychophysics*

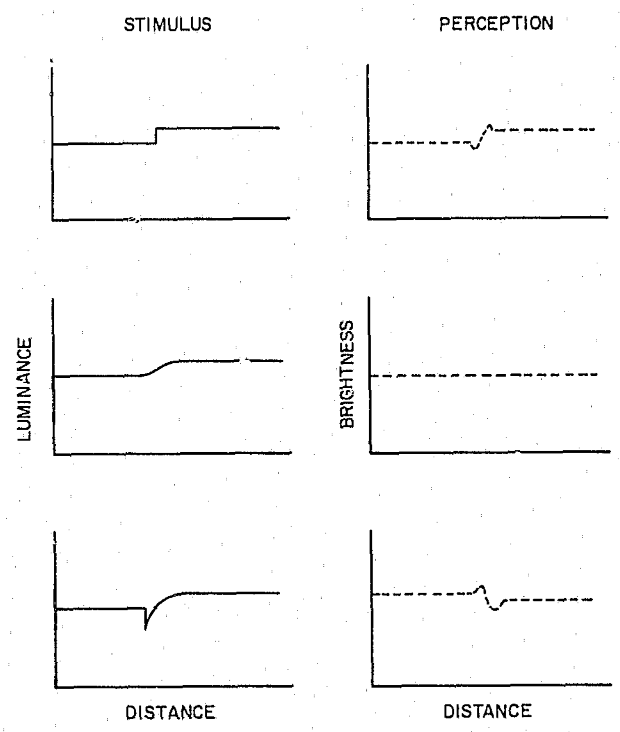
Three levels of visual task requirements have been described and differentiated according to the use to which the visual input is applied. A description of the three levels in terms of their psychophysical correlates is necessary to indicate the change in the physical counterparts, if any, associated with each level. The basic response of the visual mechanism is the discrimination of changes in luminous intensity. In detection, this discrimination between two luminance levels—contrast information—is all that is required. In recognition, "the task consists of (1) detecting the presence of a signal, then (2) assigning this signal to a category which has definitive class properties . . . Whatever the theoretical position investigators have adopted, all agree that the contour (outline) of a figure is the 'cue' or 'information' carrier for shape" [5]. Thus, for detection, a change in luminous intensity sufficient for the determination of the presence of a signal defines the total requirement, but for recognition the additional requirement of contour discrimination is imposed.

Identification differs from recognition in that the former involves the perception of details as well as contours. Identification involves the aspect of image quality that improves clarity of detail. Furthermore, as Perrin points out, "The attribute of a photograph that represents the clarity with which details are reproduced may be termed photographic definition, and it is probably determined in large measure by the sharpness of the picture . . . If, as seems likely, the most important attribute of definition is sharpness, the criterion of interest must be some property of the microdensitometer trace of the edge of the image—the graph relating the change of density at the edge to the distance across the edge" [1]. Perrin's point of view is supported by Brainard and Caum's [9] results obtained from their studies of differentiation enhancement techniques, which involved modification of the edge gradients by sharpening. They found that the enhancement technique became more effective as the subject's task became more difficult; that is, from detection to recognition to identification. For example, by sharpening the edge gradients of their images, the relative improvement in performance was 35 percent for target detection tasks, but 110 percent for identification tasks. In another comparison, the relative improvement of 17 percent obtained for recognition tasks contrasted with a 57 percent improvement for the corresponding identification tasks. Fox [5] found that recognition thresholds were slightly more affected by edge gradient changes than were detection thresholds. These findings indicate the increasing importance of image edge traces as task complexity increases, in agreement with Perrin.

The studies discussed above indicate that in the evaluation of image quality for identification, as opposed to simple detection, the edge gradient appears to be an important if not the critical factor affecting clarity of detail. In the area of psychophysiology we have further experimental studies indicating how edge gradients can importantly affect what we see. In his studies of neural interaction, Ratliff found that the sharpness of an image is to a great extent a function of the shape of the edge gradient of the image. Neural interactions accentuate the contrast at sharp spatial gradients and at discontinuities in the retinal image, resulting in a "crispening" of borders and contours [10]. This phenomenon is called the Mach band and has been extensively investigated. A comprehensive review of the experimental studies and neural models to explain this phenomenon has been made by Ratliff [11]. A good example of a similar phenomenon at the physical level is the contrast effect produced by the electrostatic field in xerography. "Wherever there is an abrupt change from a light (uncharged) area to a dark (charged) area, more powder is attracted at the edge of the dark area and less at the edge of the light area than in the uniformly dark or uniformly light areas. In order to obtain a copy with continuous tone, a halftone screen is used to break up the image so that there are changes from light to dark everywhere" [12]. As we shall see in the next study to be discussed, not only is the contour between two areas affected by the contrast between those areas, but the contour itself strongly affects the contrast.

O'Brien [13] has demonstrated that by contour manipulation the eye can be made to perceive images very different from the physical object distribution. Figure 1 is taken from O'Brien. The three (stimulus) curves to the left show the luminance distributions of the objects. The luminance differences between the left and right sides of each stimulus is above threshold, and in all cases the luminance of the left portion is less than that of the right. The edge gradients are different in all cases. The three (perception) curves to the right of the figure show the brightness distributions perceived when the object luminance distributions are those presented in the curves to the left. At the top we have the typical Mach band. A dark line is perceived on the darker half of the knife edge and a bright line is seen on the lighter half. In the middle pair, although the luminance difference is above threshold, the slowly changing edge gradient results in the perception of a homogeneous field. At the bottom is a dramatic example of how contrast can be reversed by a proper choice of the flux distribution along the edge. Physically the luminance of the left side is lower than the right side, but perceptually the left portion is brighter than the right. Perceived contrast as well as the sharpeness of contours are therefore impor-

FIGURE 1. The effect of edge gradient on perceived contrast. The luminance distribution of the object (stimulus) is given on the left and the perceived image (perception) is shown on the right.

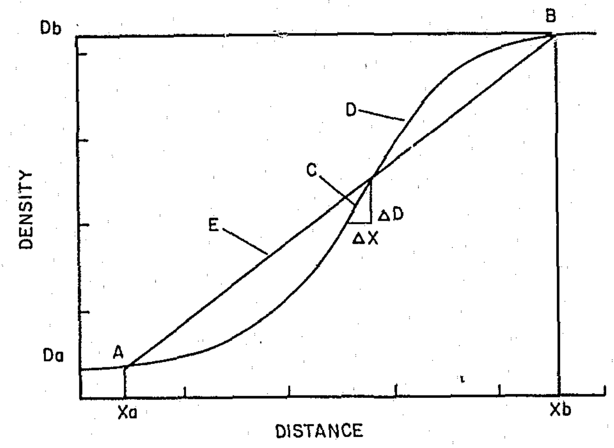


tantly influenced by the shape of the flux distribution along the boundary of an image. For this reason a physical measure that is used to describe the clarity of details should have the edge gradient as its basis.

2.3 *Edge Gradient*

When a uniformly illuminated field is cut abruptly by an adjacent uniformly dark field, we have a "knife edge." The luminance distribution along the edge separating these two fields is the edge gradient. Figure 2 is a schematic presentation of a microdensitometer trace of a knife edge. The ordinate is given in terms of density rather than transmittance or reflectance since the reponse of the eye to luminance differences is essentially logarithmic [14]. The edge gradient distance X_a to X_b is the distance perpendicular to the contour. The quantity of interest is the slope of the gradient, $\Delta D/\Delta X$. There are several ways to describe

FIGURE 2. Hypothetical microdensitometer traces of two edge gradients, E and D.



the edge gradient and its slope. Wolfe and Eisen [15] have empirically shown that the maximum gradient slope does not correlate with sharpness. They have also shown that the average gradient slope does not correlate with sharpness. In figure 2, curves E and D give the same average edge gradient. Higgins and Jones [14] found an analysis of the edge gradient that did correlate well with sharpness. They produced edge gradients from knife edges under the identical conditions used by Wolfe and Eisen. These knife edge images were microphotometered and analyzed by a measure they called acutance. Acutance correlated 0.994 with the numerical index of sharpness assigned to a series of pictures by Wolfe and Eisen.

2.4 *Actuance*

In figure 2 the distance on the sample is divided into equal increments, ΔX, taken between X_a and X_b. The slopes of the increments are squared, summed and divided by the number of increments. Mathematically,

$$G_x^2 = \Sigma(\Delta D/\Delta X)^2/N \qquad 1.$$

Higgins and Jones [14] recommend $\Delta D/\Delta X = 0.005$ where $\Delta X =$ one micrometer as the upper and lower limits to be considered in the gradient. Acutance is defined as:

$$A_c = G_x^2 (D_b - D_a) \qquad 2.$$

where D_b and D_a are the densities at X_b and X_a.

Perrin [1] suggests that sharpness is probably inversely proportional to the density scale and therefore the definition should be:

$$A_c = G_x^2/(D_b - D_a) \qquad 3.$$

In all further discussions acutance will be defined as equation (3). Other researchers have also used acutance and found a high correlation with image quality. In investigating techniques for enhancing image quality, Brainard and Caum [9] found that photo interpreters' performance is approximately a linear function of the logarithm of acutance. Thompson, [16] starting with the assumption "an image of good quality is any image which allows us to recognize and measure with high precision and accuracy an edge or boundary", determined the degree of agreement between subjective edge location by different observers. The standard deviation of four observers' means for each edge was taken as a measure of image quality. This definition of image quality correlated −0.80 with acutance, which is highly significant.

2.5 *Modulation Transfer Function*

A very popular measure of lens and system performance in use today is the modulation transfer function (MTF). As we shall see, the edge gradient is directly related to the MTF by Fourier transformation. The derivative of the edge gradient is called the line spread function (LSF). Since the first derivative is the slope of the edge gradient at x, a plot of the slope against x is the LSF. Expressed mathematically,

$$LSF = A(x) = dL(x)/dx \qquad 4.$$

and the edge gradient at x_o is

$$L(x_o) = \int_{-\alpha}^{x_o} A(x)dx \qquad 5.$$

The LSF can also be thought of as the image of a very thin slit, although the fundamental parameter of modulation transfer function is the spread of a point source rather than a line spread. But, in practical microphotometry, a split aperture is used to increase the energy available for measurement. Therefore, for purposes of this paper the integral of the point spread function in one direction, that is, the LSF, will be utilized.

The luminance distribution of the image of any object which has constant luminance in one direction can be determined by convolution of the LSF with the object distribution. Let G(x) be the luminance distribution as a function of distance x for a one-dimensional object. Take a relatively simple distribution, the cosine distribution:

$$G(x) = b_o + b_1 \cos 2\pi vx \qquad 6.$$

where b_o is the mean luminance and b_1 is the amplitude and v the spatial frequency (see figure 3). Given the luminance distribution of the object, G(x), and the LSF, A(e), the luminance of the image at point x is obtained by:

$$F(x) = \int_{-\alpha}^{\infty} A(e)G(x-e)de \qquad 7.$$

this operation being called the convolution of the object function by the line spread function [17].

For the cosine object distribution we have:

$$F(x) = b_o \int_{-\alpha}^{\infty} A(e)\, de + b_1 \int_{-\alpha}^{\infty} A(e) \cos 2\pi v(x-e)\, de \qquad 8.$$

We normalize eq (8) by dividing through by

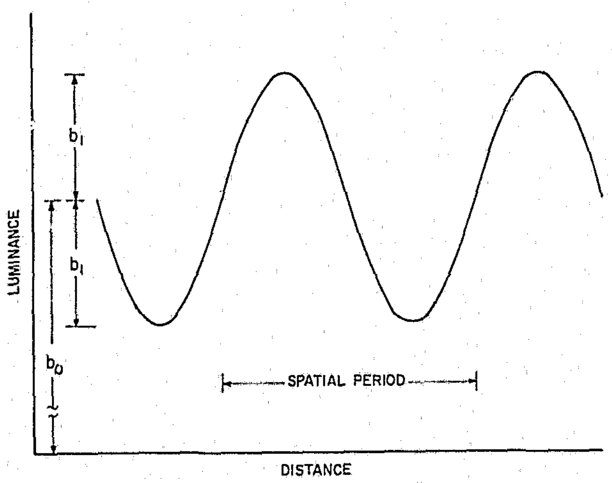


FIGURE 3. Sine-wave distribution illustrating definition of modulation $= b_1/b_o$.

$\int_{\infty}^{\infty} A(e)\, de$, the area under the LSF. The normalized equation may be rewritten as:

$$F(x) = b_o + b_1 |A(v)| \cos(2\pi vx - \phi) \qquad 9.$$

where;

$$|A(v)| = [A_c^2(v) + A_s^2(v)]^{1/2} \qquad 10.$$

is the Fourier transform of the LSF, ϕ is the phase difference between object and image, and the subscripts c and s refer to cosine and sine, respectively. The phase angle vanishes for a symmetrical line spread function, $|A(v)| = A_c(v)$; that is, the MTF and optical transfer function are the same. For purposes of this paper we will treat only the modulus of the optical transfer function. A complete derivation of eq (9) and a detailed discussion of the phase transfer function is given by Lamberts [18]. In eq (9), F(x) is a maximum when $\cos(2\pi vx) = 1$, and is at a minimum when $\cos(2\pi vx) = -1$. Therefore:

$$F(x)_{max} = b_o + b_1 |A(v)| \text{ and} \qquad 11.$$

$$F(x)_{min} = b_o - b_1 |A(v)|$$

Modulation is defined as:

$$M = \frac{L_{max} - L_{min}}{L_{max} - L_{min}} = \frac{b_1}{b_o}$$

since $L_{max} = b_o + b_1$ and $L_{min} = b_o - b_1$ (see figure 3). Given object modulation, $M = b_1/b_o$, image modulation,

$$\frac{[b_o + b_1|A(v)|] - [b_o - b_1|A(v)|]}{b_o + b_1|A(v)| + b_o - b_1|A(v)|} = \frac{b_1}{b_o}|A(v)| = M|A(v)| = M \qquad 12.$$

or the modulation transfer function,

$$MTF(v) = |A(v)| = \frac{M'}{M} \qquad 13.$$

Equation (13) states that MTF is the ratio of the modulation in the image to that in the object.

The two physical indexes discussed, acutance and modulation transfer function, will be used to describe the degradation acceptable for identification purposes. They were chosen for their direct relation to the edge gradient.

3. EXPERIMENT

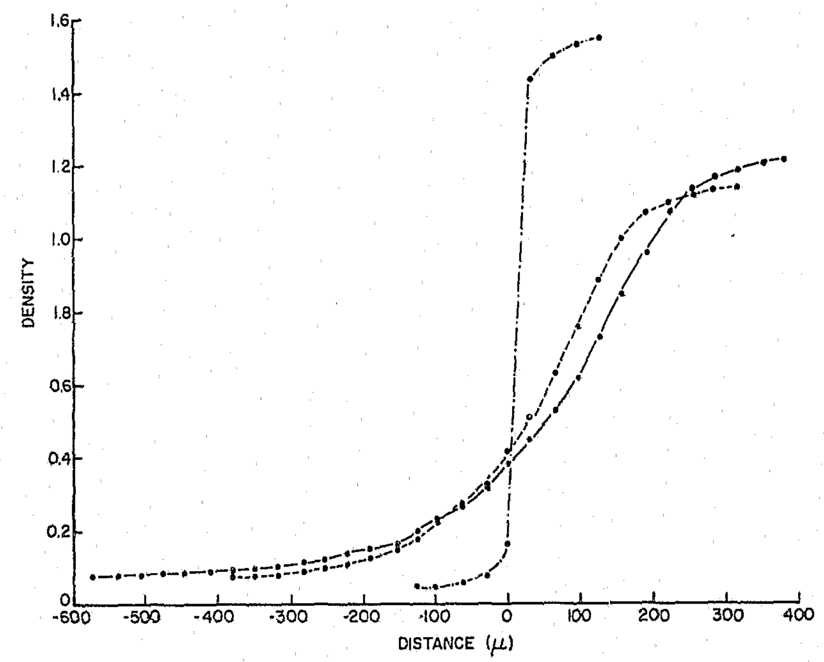
The purpose of this study was to determine the degradation tolerable in a photograph when the visual task is identification. This subjective evaluation of the adequacy of a visual target for identification purposes was obtained from a sampling of the general population. The degradation level acceptable by the average observer and a more stringent criterion, acceptable by at least 90 percent of the population, were to be determined. These two degradation levels were described in physical terms (acutance and MTF) in order that the image corresponding to the subjective evaluations could be measured directly by physical techniques.

3.1 *Stimulus*

Human faces were chosen as the identification target since they are most frequently encountered in law enforcement situations. Several aspects were considered in choosing the faces to be used in the study. 1) They had to be faces unfamiliar to the observers. 2) They should not have special features such as scars that would simplify identification. We were interested in the quality of a photograph required for identification of facial features involving clarity of details and not obvious characteristics like scars, hair style, beards, etc., which can be perceived even under highly degraded conditions. 3) The photographs had to be taken and processed under identical conditions.

The facial photographs were obtained from the local police station where facilities were available for taking of "mug" shots under standard conditions and procedures. The individuals were policemen or police recruits. Although it was felt that the probability of an observer being familiar with any one of the test faces was small, all observers were told to notify the experimenter if the individuals portrayed were familiar. All were Caucasian males with no visible scars or other distinguishing features. Front and profile shots were taken of 40 individuals. An 8 x 10 inch photograph of a knife edge (one-half black and the other half white

FIGURE 4. Microphotometry of knife edges. The dash-dot curve is the original knife edge, the dashed curve is for degradation level 9 and the solid curve is that obtained under degradation level 11.



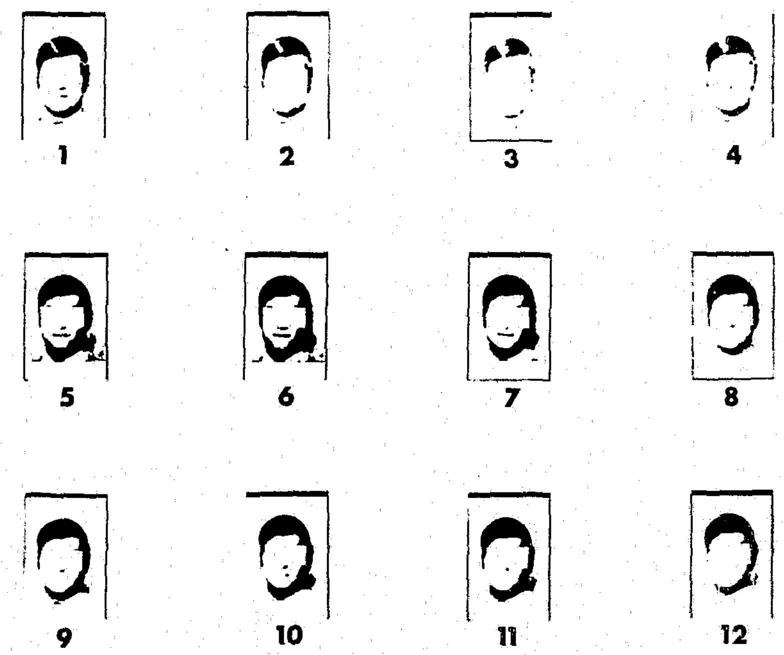
with a sharp separation between the two) was also taken under the same conditions. The edge gradient of the knife edge is given in figure 4. The 40 faces were then presented to a panel of observers and the judgments of these observers were used to obtain the ten most similar faces. One of the ten faces, randomly chosen, was degraded to 16 levels of approximately equal degradation steps. The vertical head size of the degraded photos was made one inch as recommended for Credit Union Security Requirements [19–20]. The Minimum Standards for Security Devices states: "photographic recording, monitoring or like devices capable of reproducing images . . . with sufficient clarity to facilitate (through photographs capable of being enlarged to produce a 1-inch vertical head size of persons whose images have been reproduced) the identification and apprehension of robbers or other suspicious characters" [19]. The degradation was accomplished by defocusing. The original photographic print from the police department and the copying camera were mounted on an optical bench. The camera was moved a predetermined and measured fixed distance to defocus the picture. In order to keep magnification constant, it was necessary in some cases to shift lens and film plate independently. The pictures were taken on Panatomic-X film and processed in DK-50(1–1) for five minutes at 70°F to obtain a gamma of approximately 0.8. The negatives were then contact printed on Kodak F-3 enlarging paper and dried unferrotyped.

These 16 levels of degraded images of the same individual were presented to 12 observers who judged them for subjective equality of differences in degradation. This evaluation was done by having the observers judge the degradation between adjacent levels as being: very large, large, just right, small or very small. These same observers were also asked to choose the level of degradation which they judged as being "just good enough" for identification purposes. This preliminary determination of the degradation level just good enough for identification significantly altered the degradation range. The preliminary photos were found to be overly weighted in the direction of too much degradation. The data from these 12 observers were used to make up a new set of 16 degradation levels. Of the 16 degradation levels for each face, only 12 were used in the experiment. Five of the sets were composed of degradation levels 3–14, the other five from levels 4–15. The 12 degraded photos of a given individual were pasted on the lower leaf of a two leaf folder. The photographs were numbered from 1 to 12 in increasing levels of degradation. A sample bottom leaf is presented in figure 5. The eyes were not masked in the photographs used in the study. The photograph numbers do not correspond to the degradation levels. In order to obtain the degradation level for the sample presented in figure 5, add 3 to the picture number to obtain the degradation level.

3.2 *Procedure*

The observers were employees of the National Bureau of Standards or student-faculty of the University of Maryland. Their ages ranged from 17 to 35 years; there were 55 females and 89 males, a total of 144. The subjects were all unpaid volunteers. Rather than having the subjects come to an experimental room, the experimenter went to the subject; that is, subjects were tested at locations convenient for them, which were usually their working areas. The illumination levels were approximately 60 ± 20 foot candles. When this illumination condition was not fulfilled, the subjects were asked to transfer to another location. All observers were seated and the folders were placed on the desk directly in front of them. The experiment started with the subjects reading the instructions. (A copy of the instructions is included as appendix A.) The observers were then asked to look at the large sharp photo on the top leaf and determine what level of facial degradation shown on the lower leaf was good enough for identification purposes. No restriction was put on length of viewing time and number of glances between clear and degraded photos or between degradation levels. The observers had no difficulty following the written instructions, but several subjects asked what was meant by "just good enough." The experimenter was non-committal in answering this question. The subjects were told to define just good enough as they saw fit, and that part of the experiment was to formulate this definition for identification purposes.

FIGURE 5. Example of stimulus used in determining level of degradation tolerated in image identification. The eyes were not occluded in the stimuli used in the experiment.



3.3 *Results*

There is no statistical basis for stating what percentage of the population must be satisfied with the quality of a visual image for a given purpose. This criterion is an arbitrary choice and we chose two levels, the arithmetic mean or the level that would be found acceptable to 50 percent of the population and a more stringent criterion, 90 percent acceptance. That is, in the latter case, 9 out of 10 observers agree that the quality of the picture is just good enough for identification purposes.

Table 1 gives the mean acceptance level for each of the ten photographs used in the study. The mean values for the different faces are not very different. Also included in the table is the grand mean (11.1) and the standard deviation (1.72) of all judgments. The average observer accepted degradation level 11.1 as acceptable for facial identification. That is, in order to be accepted by approximately one-half of the population, the target to be identified must have equal or higher clarity than degradation level 11.1. In order to include at least 90 percent of the population in a normal distribution a one-tailed distribution requires a standard score of 1.28. This value gives $11.1-(1.72)\ (1.28)=8.9$ as the degradation level acceptable by at least 90 percent of the population. We round these figures to 11 and 9 as the levels accepted by the average observer and 90 percent of the population, respectively.

The edge gradients for the knife edges corresponding to degradation levels 9, 11 and the master (undegraded photograph) were analyzed. The brightness spotmeter measured the luminance of a slit 2.5 x 25 micrometers on the image and the readings were continuously recorded on an x-y recorder. The grid lines on the recorder corresponded to 32 micrometers on the image plane. The edge gradients are presented in figure 4 and table 2. (For a recent discussion and review of edge gradient analysis see Swing and McCamy [22].) The master knife edge has a sharp edge and a significant change in the shape of the edge gradient is observed for degradation levels 9 and

TABLE 2. Edge gradient of knife edge degraded identically to photographs acceptable by the average observer and 90 percent of the population.

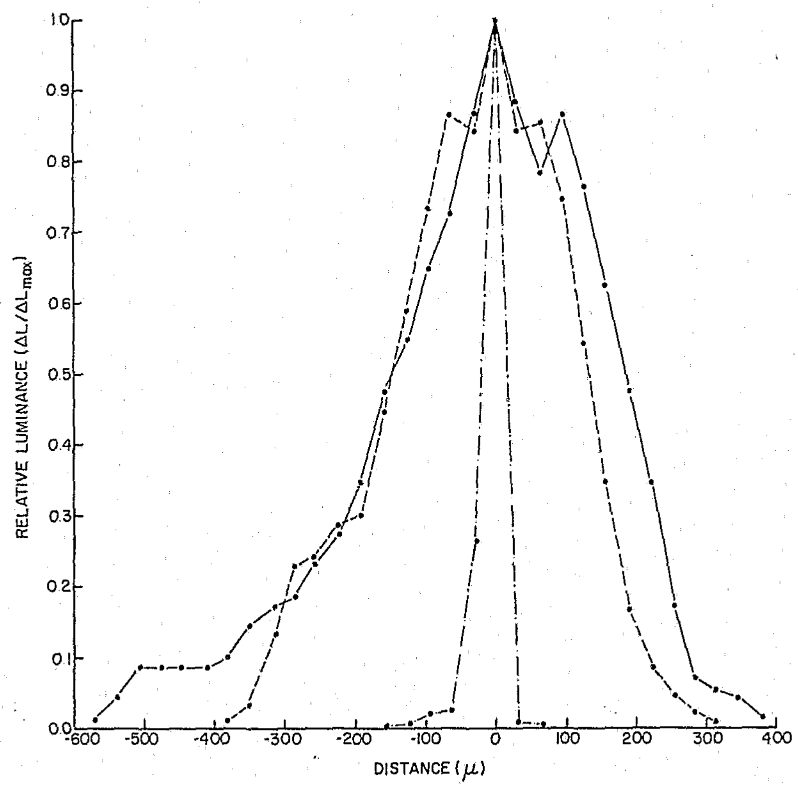
Distance (Micrometers)	Optical Density		
	Master	Mean	90% of Population
380			1.222
349		1.149	1.215
317		1.143	1.194
285		1.137	1.168
254		1.125	1.137
222		1.102	1.071
190		1.066	0.963
158		1.000	0.848
127	1.553	0.889	0.733
95	1.538	0.760	0.623
63	1.509	0.627	0.526
32	1.432	0.513	0.454
0	0.168	0.424	0.384
−32	0.070	0.328	0.317
−63	0.062	0.276	0.266
−95	0.054	0.220	0.228
−127	0.052	0.178	0.196
−158	0.051	0.148	0.171
−190		0.126	0.150
−222		0.111	0.136
−254		0.098	0.124
−285		0.087	0.115
−317		0.077	0.108
−349		0.071	0.101
−380		0.070	0.096
−412			0.092
−444			0.089
−476			0.086
−507			0.082
−538			0.079
−571			0.078

TABLE 1. The level of degradation acceptable for photographs of faces used in identification.

Picture	1	2	3	4	5	6	7	8	9	10	Mean
Mean Acceptance Level	11.4	10.8	11.3	11.5	11.6	11.4	10.6	10.5	11.2	10.9	11.1*

*Standard deviation=1.72.

FIGURE 6. Line Spread Functions for the original knife edge (dash-dot), degradation level 9 (dash) and degradation level 11 (solid).



11. $X=0$ represents that portion of the edge gradient with the largest slope.[1]

Acutance computed by eq (3) was 18.78 for the master, 4.35 for degradation level 9 and 3.39 for degradation level 11. That is, the average observer will accept for identification purposes a photograph of a face degraded as much as a knife edge degraded to give an acutance value of 3.39. But if we want 90 percent of the population to accept the photograph as just good enough for identification purposes the acutance value must be increased to 4.35.

The LSFs derived from the edge gradients are graphically shown in figure 6 and the values given in table 3. The LSF obtained in this study and in most general cases, eq (4), A(e), cannot be expressed analytically. Therefore this function must be considered point by point and the Fourier transform integration must be treated as a summation process. In mathematical notation we have;

$$MTF(v) = \frac{\sum_{i=n}^{m} A(e_i)\cos 2\pi V e_i \Delta e_i}{\sum_{i=n}^{m} A(e_i)\Delta e_i} \qquad 14.$$

The MTF derived from these spread functions using eq (14) are presented in figure 7 and table 4. In figure 7 the abscissa is given as frequency in cyc/mm and cys/deg, since in most publications on human response to sinusoidal gratings the data are generally presented as cyc/deg or in cyc/mm on the retina. The MTFs were not corrected for the slight gradient in the master knife edge (see fig. 4). Figure 7 and table 4 are interpreted as indicating the modulation ratio required between image and object when the goal is an image acceptable for identification purposes. The solid curve in figure 7 is the modulation ratio

[1] The edge gradients from the degraded knife edges were not corrected for the small deviations from a perfect step function. When applied to the MTF, the corrections for the spatial frequencies of interest are negligible.

accepted by the average observer. If we want to include not only the average observer, but 90 percent of the population, the modulation ratio required is that level indicated by the dashed curve. For example, the modulation ratio of a sine wave of frequency 7.98 cyc/deg should be 0.4811 for picture clarity to be acceptable for identification purposes by 9 out of 10 observers. When the spatial frequency is increased to 10.11 cyc/deg the modulation ratio required is 0.3148. As will be shown in discussions to follow, the frequency and MTF relations given in figure 7 and table 4 apply only to a vertical head length of 4.8 deg.

TABLE 3. The Line Spread Function obtained from edge gradient of knife edge degraded identically to photographs acceptable by the average observer and 90 percent of the population.

X (Micrometers)	Reflectance		
	Master	Mean	90% of Population
380			0.014
349			0.043
317		0.012	0.058
285		0.024	0.072
254		0.048	0.174
222		0.084	0.348
190		0.169	0.478
158	.003	0.349	0.623
127	.008	0.542	0.768
95	.023	0.747	0.870
63	.025	0.855	0.783
32	.268	0.843	0.884
0	1.000	1.00	1.00
– 32	.009	0.843	0.870
– 63	.003	0.867	0.725
– 95	.002	0.735	0.652
–127		0.590	0.551
–158		0.446	0.478
–190		0.301	0.348
–222		0.289	0.275
–254		0.241	0.232
–285		0.229	0.188
–317		0.133	0.174
–349		0.036	0.145
–380		0.012	0.101
–412			0.087
–444			0.087
–476			0.087
–507			0.087
–538			0.043
–571			0.014

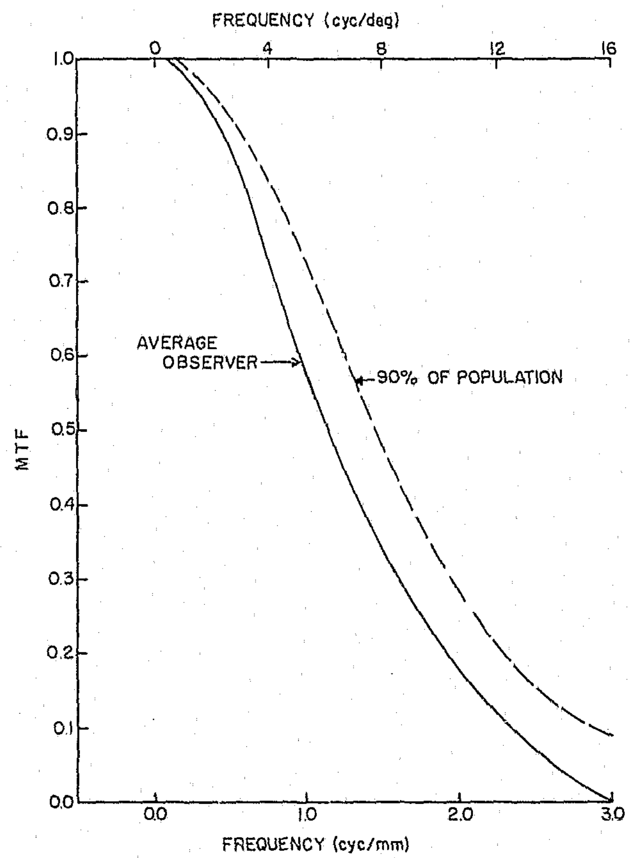
TABLE 4. Modulation Transfer Function acceptable for identification purposes by the average observer and 90 percent of the observers.

Frequency		Modulation Transfer Function	
cyc/mm	cyc/deg	Mean	90% of Population
.1	.532	.9942	.9967
.2	1.06	.9770	.9870
.3	1.60	.9492	.9710
.4	2.13	.9119	.9491
.5	2.66	.8666	.9216
.6	3.19	.8151	.8890
.7	3.72	.7591	.8520
.8	4.26	.7006	.8112
.9	4.79	.6414	.7674
1.0	5.32	.5829	.7212
1.1	5.85	.5267	.6734
1.2	6.38	.4735	.6248
1.3	6.92	.4242	.5761
1.4	7.45	.3789	.5280
1.5	7.98	.3378	.4811
1.6	8.51	.3005	.4359
1.7	9.04	.2666	.3928
1.8	9.57	.2357	.3524
1.9	10.11	.2073	.3148
2.0	10.64	.1807	.2803
2.1	11.17	.1558	.2489
2.2	11.70	.1321	.2206
2.3	12.23	.1096	.1955
2.4	12.77	.0883	.1733
2.5	13.30	.0684	.1539
2.6	13.83	.0502	.1369
2.7	14.36	.0338	.1222
2.8	14.89	.0196	.1094
2.9	15.43	.0078	.0983
3.0	15.96	–.0015	.0884
3.1	16.49	–.0082	.0794
3.2	17.02	–.0125	.0712
3.3	17.55	–.0147	.0635
3.4	18.09	–.0150	.0561
3.5	18.62	–.0139	.0487

3.4 *Discussion*

Birch [8], in reviewing the literature on image quality, concludes that there does not appear to be one index that will fully describe image quality. However, he points out that if a required characteristic is explicitly stated, such as detail of content, frequently a well correlated physical counterpart can be found. In the introduction it was argued that image quality as related to identification involves clarity of detail and may be more specifically defined as the sharpness of the image. Furthermore, the sharpness of an image is most directly influenced by the edge gradient. The measure of image quality in which the edge gradient is directly and fully utilized is acutance.

FIGURE 7. The modulation ratio required between object and image when the intended use is image identification.



The MTF may be derived from the edge gradient, but as stated by Brock, [22] "There is no need to go all the way to the extraction of the MTF with inevitable errors; since all the information resides in the edge image itself, it seems logical to keep as near as possible to this." If, as discussed in this paper, image quality for identification is importantly dependent on the edge gradient, any measure that is linked only to the edge gradient will not be as powerful as one directly utilizing the primary physical variable. Yet, the increasing use of MTF in lens and system performance makes it highly desirable to utilize it as a common measure. Furthermore, a subjective image quality measure that can be directly cascaded with measures used to evaluate image producing systems is preferred. From a practical standpoint, the observation that spatial frequencies over a limited range only need be considered, argues strongly for MTF. The maximum and minimum for sine-wave gratings of only ten selected frequencies need be measured. The measurement of maximum and minimum is a relatively simple task and is less susceptible to measurement error than edge gradient analysis. Since simplicity of measurement is an important criterion in the choice of a physical measurement technique, this argues for the use of MTF. Based on the arguments stated above, the writer recommends the use of the MTF index for image identification acceptance level.

The angular subtense used in this study was 4.8°, a one inch (25.4 mm) vertical head height observed at a distance of 12 in. (304.8 mm). The retinal image of an object viewed at 12 in. will be twice as large as that viewed at 24 in. Similarly, if the object size is doubled and viewing distance kept constant, the retinal image will be doubled. The MTF presented in figure 7 applies to a vertical head size of 4.8°. The MTF for angular subtenses other than that investigated in this study have been calculated.

The spatial frequency relative to the image dimension is the pertinent parameter, that is, (cyc/deg) (deg/vertical head length) = cyc/vertical head length. The modulation ratio required for acceptance by 90 percent of the population is presented in figure 8 and table 5 as MTF against spatial frequency in cyc/vertical head height. For example, if the vertical head height is divided into 23 cycles, the modulation ratio required is 0.77. If there are 33 cyc of a sine-wave across the vertical head height of the face, then the ratio must be 0.58; that is, (image modulation/object modulation) must be 0.58 if we want the target to be acceptable by 90 percent of the population. No matter what the absolute dimension of the image, when the vertical head height is 33 cycles, the MTF must be 0.58.

The illumination level used in this study was 60 ± 20 foot candles. Visual performance is dependent on luminance; therefore, in the strictest sense, the results of this study apply only to

FIGURE 8. The modulation ratio required for identification of a face as a function of number of cycles per vertical head height.

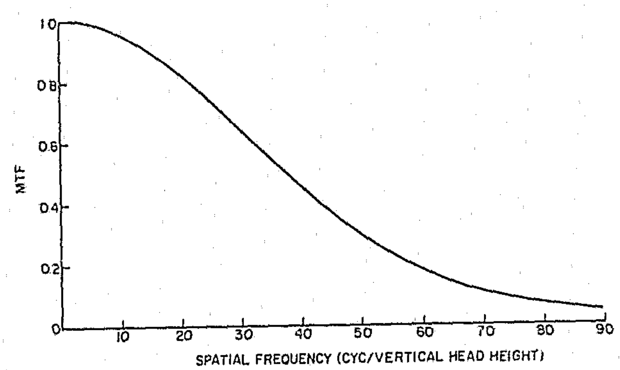


TABLE 5. Modulation ratio of sine-wave targets required for image identification, where spatial frequency is given relative to vertical head size.

Cyc per Vertical Head Length	Modulation Transfer Function 90 Percent of Population
2.55	0.9967
5.09	0.9870
7.68	0.9710
10.22	0.9491
12.77	0.9216
15.31	0.8890
17.86	0.8520
20.45	0.8112
22.99	0.7674
25.54	0.7212
28.08	0.6734
30.62	0.6248
33.22	0.5761
35.76	0.5280
38.30	0.4811
40.85	0.4359
43.39	0.3928
45.94	0.3524
48.53	0.3148
51.07	0.2803
53.62	0.2489
56.16	0.2206
58.70	0.1955
61.30	0.1733
63.84	0.1539
66.38	0.1369
68.93	0.1222
71.47	0.1094
74.06	0.0983
76.61	0.0884
79.15	0.0794
81.70	0.0712
84.24	0.0635
86.83	0.0561
89.38	0.0487

60 ± 20 foot candles. The determination of the luminance level correction is beyond the scope of this study.

The angular subtense (cyc/deg) for an image intensifier with a 65 mm focal length objective is given in table 6 (see columns 1 and 2). The cyc/mm is measured at the scope face. They were converted to cyc/vertical head length and MTF values obtained from table 5 by linear interpolation. Figure 9 and table 6 present the results of the above interpolation. The circles and crosses depict contrast (modulation) data obtained from two image intensifiers with 65 mm focal length objectives using a three bar crenelate target with a contrast of 0.95.[2] Figure 9 indicates that one of the image intensifiers will perform satisfactorily

[2] No correction has been made for the fact that the targets were not sinusoidal targets with a modulation of 1.0.

FIGURE 9. The modulation ratio required for identification of a face. The parameter 2-8 deg is the angular subtense of the vertical head height.

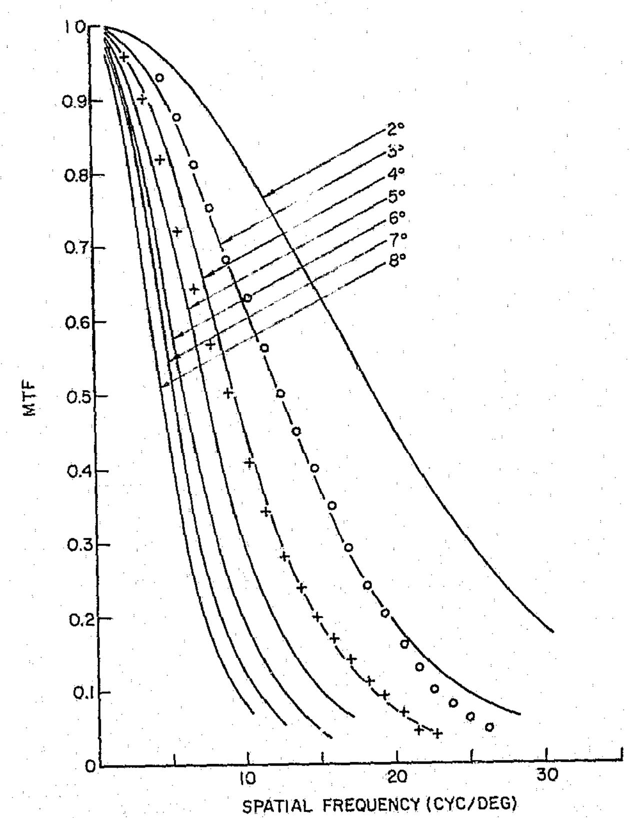


TABLE 6. Modulation ratio of sine-wave targets required for image identification as a function of spatial frequency. The parameter, vertical head height, is directly related to observing distance or magnification. The spatial frequency in cyc/deg applies to an image intensifier with a 65 mm focal length objective.

Spatial Frequency		Vertical Head Height in Degrees						
cyc/mm	cyc/deg	2	3	4	5	6	7	8
1	1.13		0.9935	0.9892	0.9849	0.9764	0.9690	0.9593
2	2.27	0.9891	0.9762	0.9590	0.9369	0.9107	0.8806	0.8472
3	3.40	0.9762	0.9493	0.9110	0.8645	0.8114	0.7065	0.6902
4	4.54	0.9590	0.9107	0.8472	0.7725	0.6894	0.6026	0.5177
5	5.67	0.9370	0.8644	0.7728	0.6682	0.5610	0.4565	0.3617
6	6.81	0.9107	0.8109	0.6894	0.5604	0.4357	0.3270	0.2394
7	7.94	0.8807	0.7524	0.6030	0.4563	0.3274	0.2271	0.1564
8	9.08	0.8472	0.6894	0.5177	0.3610	0.2294	0.1561	0.1043
9	10.21	0.8110	0.6246	0.4361	0.2808	0.1732	0.1094	0.0713
10	11.34	0.7728	0.5610	0.3617	0.2153	0.1273	0.0787	
11	12.48	0.7317	0.4917	0.2960	0.1649	0.0951	0.0546	
12	13.61	0.6898	0.4363	0.2398	0.1272	0.0714		
13	14.75	0.6463	0.3792	0.1929	0.0994	0.0513		
14	15.88	0.6030	0.3274	0.1564	0.0789			
15	17.02	0.5606	0.2806	0.1270	0.0610			
16	18.15	0.5181	0.2397	0.1044				
17	19.28	0.4765	0.2041	0.0866				
18	20.42	0.4361	0.1732	0.0713				
19	21.55	0.3979	0.1485	0.0578				
20	22.69	0.3614	0.1271					
21	23.82	0.3274	0.1096					
22	24.96	0.2960	0.0951					
23	26.09	0.2666	0.0826					
24	27.23	0.2396	0.0713					
25	28.36	0.2151	0.0611					
26	29.49	0.1931						
27	30.63	0.1732						

for image sizes as small as 3 in angular subtense, but that the second is limited to angular sizes of 4° or larger. The similarity of the image intensifier and image identification functions support the use of this measure as the index for identification.

Following the suggestion of Granger and Cuprey [23], a single number index may be obtained by integrating the MTF over a limited frequency band. They found that an MTF based index of image quality correlated highly with film print quality rank. The quality of the prints were varied by simple defocus, contrast loss, Gaussian degradation, simulated central obstruction of lens aperture, film adjacency effect and astigmatism. In instructing their subjects they carefully avoided using specific terminology; for example, sharpness and resolution. They were interested in overall judgments of quality and thus they used a paired comparison technique in which the subjects were asked "If you could keep only one picture, which picture would you prefer to keep?," or "Which picture was taken by the better camera?" The proposed Subjective Quality Factor (SQF) weights the MTF curve by including only that portion of the spatial frequency spectrum important to the human eye, 3.0 to 11.9 cyc/deg.[3] As can be seen in figure 7, this range includes the main portion of the frequency spectrum of interest for identification of a face subtending 5°, the approximate image size used in this study. But for smaller angular subtenses the frequency range has to increase in order to include the relevant frequencies. Furthermore, in identification the sharpness of the image is assumed to be the most important parameter. Therefore, when image quality is restricted to

[3] In converting from line pairs/mm at the retina to cyc/deg, 17 mm was used as the interocular distance of the eye.

identification, the frequency band should be enlarged to include higher frequencies than those suggested by Granger and Cuprey, the recommended range being 3–20 cyc/deg.

From figure 9 it can be noted that a summation of MTF over the frequency range 3–20 cyc/deg will give an index increasing with decreasing angular subtense of vertical head length and will cover frequencies relevant to image identification performance. A 2° vertical head height will cover approximately ⅛ the diameter of a 18 mm scope face when the image intensifier has a 65 mm focal length objective. There is no empirical evidence to indicate which single frequency, if any, is more effective in image identification. Therefore, a single number index based on single frequency, although highly desirable, has no empirical justification.

4. RECOMMENDATION

From the results of this study and the sample contrast measurements obtained from two image intensifiers, we recommend the use of the MTF vs cyc/deg function as the acceptance index for identification. A step by step procedure and examples demonstrating the computation techniques for this index are given in appendix B.

For a single number index we recommend summation of MTF over the frequency range 3–20 cyc/deg. In mathematical notation,

$$\text{identification quality index} = \sum_{i=3,\ \text{step } i=2}^{19} MTF_i\,(H);$$

where the cyc/deg is taken at every other degree and H represents the parameter head size or viewing distance.

APPENDIX A. INSTRUCTIONS

This is a study to determine the quality of an image required to recognize a face from a photograph.

Please open the folder in front of you. Do it now.

Look at the bottom leaf with the twelve facial photographs. As you can see, photograph 1 is the clearest and photograph 12 the least clear, with the other numbers representing intermediate stops. Choose the photograph you feel is JUST GOOD ENOUGH to identify the individual shown on the top leaf. When you have made your decision give the experimenter the photograph number. The photographs can be viewed in any sequence and you may take as long as you need.

You will be asked to follow the same procedure for 9 additional faces. Are there any questions? If not you may begin.

APPENDIX B.

Step by step procedure to obtain MTF for given angular subtense of vertical head height and spatial frequency in cyc/u, where u is the unit measure. Transform line pairs/u or cyc/u to angular measure, cyc/deg.

$$\text{cyc/deg} = fd/57.3 \qquad \text{1b}$$

where: $f = \text{cyc/u}$,
d = viewing distance or focal length, and
f and d are in the same units.

Example: $f = 2$ line pairs/mm and $d = 65$ mm, then $2(65)/57.3 = 2.27$ cyc/deg.

Compute angular subtense of vertical head height (H).

$$\text{deg/H} = (57.3H)/d \qquad \text{2b}$$

where H and d are in the same units.

Example: $H = 25$ cm and $d = 358$ cm, then

$$(57.3)(25)/358 = 4 \text{ deg, or}$$

for H measured on the face of the scope $= 4.5$ mm with objective focal length $= 65$ mm;

$$(57.3)(4.5)/65 = 3.97 \text{ deg.}$$
$$\text{cyc/H} = (\text{deg/H})(\text{cyc/deg})$$

Convert into cycles per vertical head height in angular notation.

$$\text{cyc/H} = (\text{deg/H})(\text{cyc/deg}) \qquad \text{3b}$$

Example: cyc/deg $= 2.27$ and deg/H $= 4$, then

$$2.27(4) = 9.08 \text{ cyc/H.}$$

From Table 5 find MTF by linear interpolation.

APPENDIX C. REFERENCES

1. Perrin, F. H., "Methods of Appraising Photographic Systems Part I—Historical Review," J. Society Motion Picture Television Engineers, Vol. 69, pp. 151–156, 1960.
2. Stulz, K., and Zweig, H., "Roles of Sharpness and Graininess in Photographic Quality and Definition," J. Optical Society of America, Vol. 52, pp. 45–49, 1962.
3. Shack, R., "A Proposed Approach to Image Evaluation," Optical Image Evaluation, National Bureau of Standards Circular 526, pp. 275–284, 1951.
4. Higgins, G. C., and Perrin, F. H., "Evaluation of Optical Images," Photographic Science & Engineering, Vol. 2, pp. 66–76, 1958.
5. Fox, W. R., "Visual Discrimination as a Function of Stimulus Size, Shape and Edge-Gradient," Form Discrimination as Related to Military Problems, Publication 561, pp. 168–175, (J. W. Wulfeck and J. H. Taylor, editors, National Academy of Sciences-National Research Council, 1957).
6. Woodworth, R. S., and Schlosberg, H., Experimental Psychology (Henry Bolt and Company, New York, 1954).
7. Harmon, L. D., "Some Aspects of Recognition of Human Faces," Pattern Recognition in Biological and Technical Systems, pp. 196–219 (Springer-Verlag, New York, 1971).
8. Birch, K. G., "A Survey of OTF Based Criteria Used in the Specification of Image Quality," (National Physical Laboratory, Teddington, England, 1969); NTIS No. N70 21982.
9. Brainard, P. W., and Caum, K. B., "Evaluation of an Image Quality Enhancement Technique," AMRL-TR-65-143, (Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, September 1965); NTIS No. AD624470.
10. Ratliff, F., "Inhibitory Interaction and the Detection and Enhancement of Contours," Sensory Communications (W. A. Rosenblith, editor, MIT Press, Cambridge, Massachusetts, 1961).
11. Ratliff, F., Mach Bands: Annotative Studies on Neural Networks in the Retina (Holden-Day, Inc., San Francisco, 1965).
12. Ratliff, F., "Contour and Contrast," Proceedings of the American Philosophical Society, Vol. 115, pp. 150–163, 1971.
13. O'Brien, V., "Contour Perception, Illusion and Reality," J. Optical Society of America, Vol. 48, pp. 112–119, 1958.
14. Higgins, G. C., and Jones, L. A., "The Nature and Evaluation of the Sharpness of Photographic Images," J. Society Motion Picture Television Engineers, Vol. 58, pp. 277–290, 1952.
15. Wolfe, R. N., and Eisen, F. C., "Psychometric Evaluation of the Sharpness of Photographic Reproductions," J. Optical Society of America, Vol. 43, pp. 914–922, 1953.
16. Thompson, L. G., "The Use of Edges of Photographic Images as Specifiers of Image Quality," ETL-CR-72-15, (U.S. Army Engineering Topographic Laboratory, Ft. Belvoir, Virginia, November 1972); NTIS No. AD753947.
17. Bracewell, R., The Fourier Transform and Its Application (McGraw-Hill Book Co., New York, 1965).
18. Lamberts, R. L., "Relationship Between the Sine Wave Response and the Distribution of Energy in the Optical Image of a Line," J. Optical Society of America, Vol. 48, pp. 490–495, 1958.
19. "Minimum Standards for Security Devices,"

Federal Register, Section 748.9, Vol. 36, No. 109, p. 10912, June 5, 1971.
20. Meltzer, R. J., "Optimum Utilization of Bank Protection and Surveillance Cameras," Police, pp. 6–10, May 1972.
21. Swing, R. E., and McCamy, G. S., "Edge Gradient Analysis: A Critique," National Bureau of Standards Report 10 119, October 1969.
22. Brock, G. C., "A Review of Current Image-Evaluation Techniques," J. Photographic Science, Vol. 16, pp. 241–249, 1968.
23. Granger, E. M., and Cupery, K. N., "An Optical Merit Function (SQF) which Correlates with Subjective Image Judgments," Photographic Science & Engineering, Vol. 15, pp. 221–230, 1972.

☆U.S. GOVERNMENT PRINTING OFFICE: 1974 O—549-901

END