

G. P. DE LOOR, A. A. JURRIENS,
*Physics Laboratory RVO/TNO
The Hague, The Netherlands*

W. J. M. LEVELT*, AND J. P. VAN DE GEER
*Institute for Perception RVO/TNO
Soesterberg, The Netherlands*

Line-Scan Imagery Interpretation

Tests show how PI performance deteriorates with image degradation.

INTRODUCTION

IN MODERN RECONNAISSANCE equipment line-scan methods are used. They offer pictures built up of contiguous lines (such as on a TV-screen). The extraction of useful and significant information from such pictures still requires a human observer. This raises questions as to:

1. The design of sensors, especially with regard to presentation of the information obtained, and

tively large part of the following study is therefore devoted to the definition of the stimulus material used in the experiments.

Two main problems, both pertinent to the development of line-scan equipment, are studied in this experiment: (1) What is an acceptable degree of resolution? and (2) Are there any advantages of stereo-presentation?

MATERIAL

A proper design of the experiment made it necessary to have stereo-photographs which

ABSTRACT: An image interpretation experiment involved the interpretability of line-scan pictures. Care was taken to use stimulus material fabricated with equipment whose transfer characteristics were well defined. System parameters as used in the experiment are given by the Modulation Transfer Function. Stimulus material thus obtained was presented to trained photo interpreters for various tasks (detection, discrimination, depth estimation). Results were analyzed as a function of two experimental variables: level of degradation, and stereo versus non-stereo presentation. Curves show the deterioration of PI accuracy with increasing degradation. Stereo presentation proved to be of very limited usefulness: it seems to increase accuracy only for typical height estimation tasks in conjunction with moderate degree of degradation. This limited advantage of stereo presentation is obtained at a cost of an overall increase in time spent at the tasks.

2. The training of photo-interpreters (PI's) in reading the unusual material presented to them.

An experiment was therefore conducted to probe the effect of image degradation, often inherent in line-scan imagery, on the extraction of information by PI's from TV-type imagery. An experiment of this type, however, can be successful and complete only if the photographic material used in the tests is clearly and unambiguously defined. A rela-

fulfill a number of requirements. First the basic photographic material should be perfectly defined as to content, as to physical parameters (height, stereo-base, etc.) and as to the transmission characteristics of the whole system. Secondly, the content has to allow for all sorts of PI-questions in systematic variation. For these reasons it was impossible to use real life aerial photographs. The mere requirement of a constant stereo-base is already prohibitive in this respect. It was therefore decided to use photographs made of a small-scale model.

* Now at University of Groningen, The Netherlands.

THE MODEL

The model was 7.5×24 feet, and was carefully built at a scale 1:86. The scene contains a small village with station, a number of factories surrounded by open terrain, a canal, a cargo station. Various movable objects could be placed in this scene, such as soliders and civilians; cars, military and civil lorries; railway waggons, lorries and engines; tanks, rockets, machine guns and other artillery; heaps of sand of different heights; factory chimneys and spruce trees of varying heights; interchangeable viaducts and bridges; a number of changeable buildings, etc. To have an opportunity to compare the detectability of objects in various contexts, four arrangements of the whole scene were made by varying the location and kind of the movable objects.

PHOTOGRAPHIC PROCEDURE

All photographs of the model were made under well defined conditions with a Rollei-cord camera ($1:3.5; f=75$ mm). This camera was thoroughly tested beforehand and test patterns were also included in the model for later evaluation of the photographic procedure.

A flashlight was attached close to the camera lens in order to obtain shadowless pictures. Camera and flashlight were mounted above the model at a height 8 ft. 10 in. For a scale of 1:86 this corresponds to an altitude of 760 ft. The stereobase of 9.8 inches corresponds to a real distance of 70 feet.

All these photographs were then degraded to line-scan pictures with a flying-spot-scanner system. (Figure 1.) Such a system is a good representation—at least for the experiment to be described—of an optical line-scan system for aerial reconnaissance and it gives the experimenter more flexibility and the

possibility of introducing degradation in well determined steps.

THE FLYING-SPOT-SCANNER SYSTEM

The system of Figure 1 consists of a cathode-ray tube and a lens. On the screen of the cathode-ray tube a raster is written by the electron beam. This raster is projected on the negative under test by the lens. This negative is thus scanned line after line. The light transmitted through the negative falls on a photomultiplier tube after passing a condenser system. This light is thus modulated in intensity by the negative. The photomultiplier gives an electrical signal proportional to this luminous intensity. This signal is amplified and fed to a TV picture tube. The number of amplification steps is such that a positive image is obtained on the screen of this tube. This positive image was photographed.

A major feature of this type of simulator is its ability to degrade image detail in well determined steps and the possibility to display/images with any arbitrarily chosen signal noise ratio. This is possible with the use of a noise generator giving a signal which can be added linearly to the video signal.

The system of Figure 1 is a 625-line system using 2:1 interlaced scanning with a frame repetition rate of 25 frames/sec. Synchronisation between the rasters of the flying-spot-scanner and the TV monitor tube is obtained with a specially designed pulse generator.

The system is considered as defined when the modulation transfer function (MTF) has been determined (Elias et al, 1952; Hempenius, 1964; Schade, 1951–1955). The MTF gives the frequency response of the system to the energy spectrum at its input. The system under consideration can be treated in three parts. A division is made in parts where

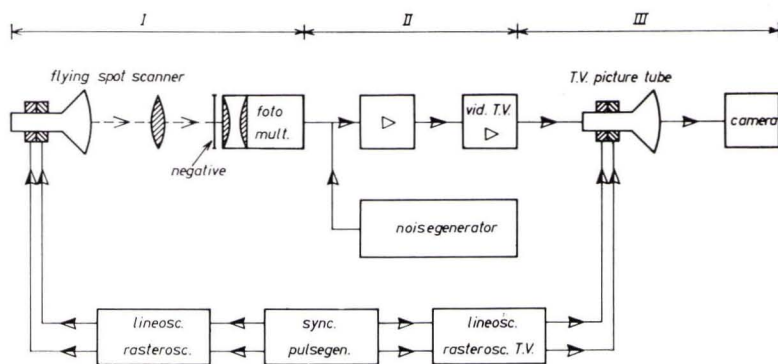


FIG. 1. Flying-spot scanner system.

electro-optical transformations take place and a purely electronic part.

- I. *Negative (original)*. Output photomultiplier tube. The modulation in luminous intensity caused by the negative is transformed into an electrical signal.
- II. *Output photomultiplier*. Cathode of TV-picture tube. The electrical signal obtained under I is amplified.
- III. *Cathode TV-picture tube*. Film in recording camera. The electrical signal is again transformed into a spatial intensity distribution, which distribution is recorded on film. The MTF of the whole system is the product of the MTF's of the three parts.

Comparable quantities are intensity and voltage. It was found that the photographic process used for the recording of the final image introduces no further degradation in image quality. This means that, in fact, we can cut stage III at the screen of the TV-picture tube, because the camera, etc., does not influence final quality.

The speed of the light spot of the flying-spot scanner over the negative determines the parameter for the transformation of the spatial intensity distribution into a distribution in time. This speed can be determined (given the line frequency) by measuring the width of the raster as projected on the negative, or by comparison of the width of the testpattern on the (input) negative with the electrical signal on an oscilloscope. The intensity distribution of input and output signal was determined by measuring the density (along a line) of the original negative and the output negative with the aid of a microdensitometer.

The following testpatterns were used:

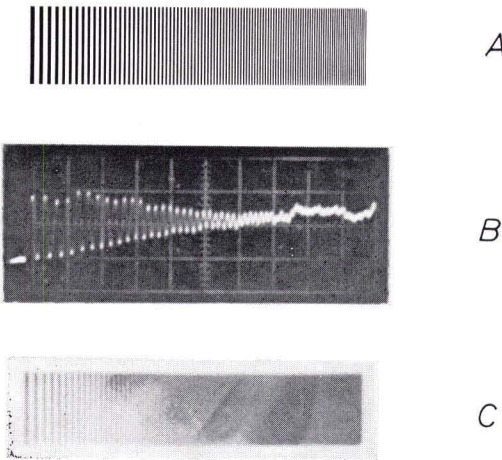


FIG. 2. a. Testpattern, b. Output of Part II, c. Reproduction on TV-screen.

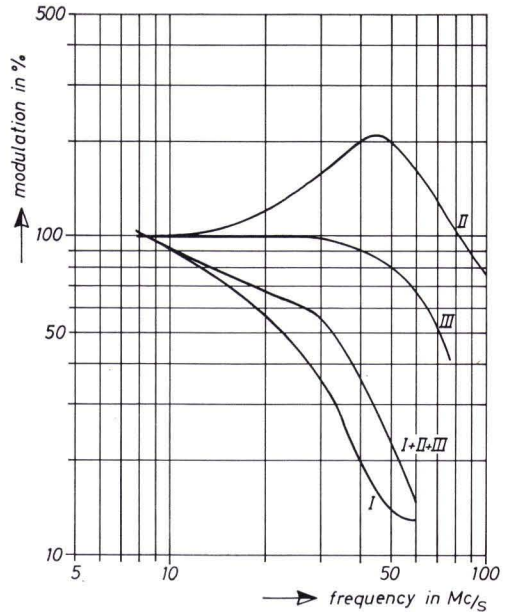


FIG. 3. Transfer characteristics of the system.

- A sine-wave pattern, in which the density varies sinusoidally. The frequency increases to one side.
- A bar pattern.
- Three bar patterns.

As an example we include in Figure 2 the bar pattern, the electrical signal caused by this pattern at the end of part II of the system, and the final image as obtained by photographing the TV-picture tube.

Figure 3 gives the MTF for the three parts of the system and that of the system as a whole.

The system is made such that resolution is insofar as possible equal in both directions: along a line and at right angles with it. This was tested by placing the test patterns in both directions. Figure 3, however, gives the MTF along a line. Examination of this figure shows that part I is limiting the system performance. The reason is the long decay time of the phosphor of the flying-spot-scanner tube.

The above analysis can be applied only if the system can be considered as linear. Although with some adaptations this analysis is also applicable to systems containing a certain amount of non-linearity (Langler and Muller, 1963), the basic theory, in principle, can only be used with linear systems.

Linearity was therefore tested with the aid of a photographic step wedge. Reproduction of this wedge by the system should give an identical wedge at the output. This proved to be the case for a density variation of 0.85 D,

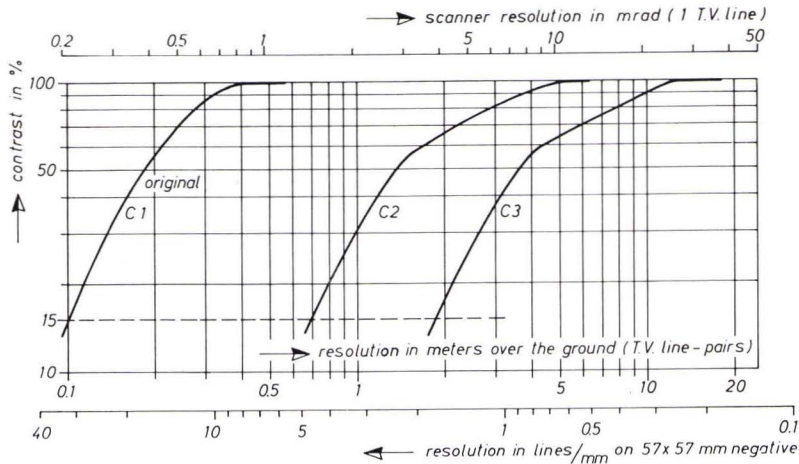


FIG. 4. Modulation Transfer Function of total system.

which means an intensity range of about 1 to 9. The limiting component is here the screen of the TV-picture tube.

PREPARATION AND QUALITY OF THE STIMULUS MATERIAL

Knowing the MTF of the camera and its place above the model we now can give resolution not only in lines/mm over the negative, but also in meters over the ground or in milliradians angular resolution. This is done in Figure 4 with curve C-1, giving the MTF of the camera, which produced the original negative, relative to a series of system parameters.

The degraded (line-scan) pictures were obtained as follows. First the setting of part I of the flying-spot-scanner system (Fig. 1) was made such that the whole negative was illuminated by the raster. This gives the best reproduction for the whole negative as possible with this system. The MTF of the whole system is then given by curve C-2 in Figure 4. For the second degradation only a part of the raster is imaged on the negative. This gives curve C-3 in Figure 4. Graph 4 gives resolution also in lines/mm, using 2.3×2.3-inch negatives for the reproduction.

Figure 5 gives an example of the images finally obtained.

The definition of image quality as described is done by giving the MTF of the system for the negatives finally obtained, as well for the original as for the reproductions C-2 and C-3. For the actual tests these negatives were printed. This printing process was not controlled. For the degraded imagery this plays no role, because the photographic process is of much higher quality than the flying-spot-scanner system. For the original, however,

(contact prints) we must expect a response somewhat worse than given by curve C-1 in Figure 4.

Because of the limited linearity, we take resolution at the 15 percent contrast level as limiting resolution. This choice is in fact a bit arbitrary and conventional, but we need a single figure per degradation to be used as reference along one of the axes of the figures in the following sections. The resulting figures for resolution are given in Table 1.

EXPERIMENTAL DESIGN

THE 3-D SYSTEM

Photo interpretation tasks were classified into three types:

Detection tasks. The PI has to look for a known object (indicated by means of a picture, a description or a name), the location of which is not given. For example, the PI is given separate photographs of a lorry; he is told that such a lorry is contained in the scene and asked to indicate the location of it by specifying coordinates.

Discrimination tasks. The location of the target is given. The P.I. has to identify the object (often among a number of alternatives) or to specify its orientations. For instance, the

TABLE 1. LIMITING RESOLUTION AT 15 PERCENT CONTRAST LEVEL

Curve in fig. 4	Over ground	l/mm on 57×57 mm negative	Angular res. (1 TV-Line)
C-1	0.1 m or 4"	33	0.2 mrad.
C-2	0.69 m or 27"	4.8	1.4 mrad.
C-3	1.85 m or 70"	1.9	3.8 mrad.

PI is given a separate picture of five different kinds of trucks; he is asked to say which of these is present at a given location of the scene.

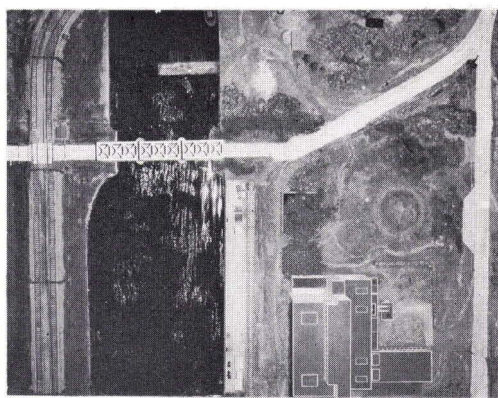
Depth tasks. Logically these tasks are only a variety of the discrimination type, but for the purpose of this study, specific depth discrimination problems were treated separately, in order to obtain data about the usefulness of stereo information. And for the same reason, non-stereo depth cues were eliminated in these tasks. Our depth tasks are therefore restricted to targets which have no characteristic height of their own. For example, the PI is asked to indicate the order of height of five trees.

Even then, it is not precluded that a PI attempts an answer on the basis of non-stereoscopic material. Our argument however is: where previous studies have denied the usefulness of stereoscopic presentation, our depth tasks allow a maximum opportunity for stereo presentation to reveal its effects. If stereo presentation has no effect here, it can no longer be assumed to have any usefulness at all within the range of military PI-tasks (with the possible exception that stereoscopy might be used for comparative cover, i.e., the detection of changes in the scene if two pictures are given which are taken successively, with a certain time interval). Of course, the usefulness of stereoscopic material for photogrammetry or various geological purposes falls outside the scope of this discussion.

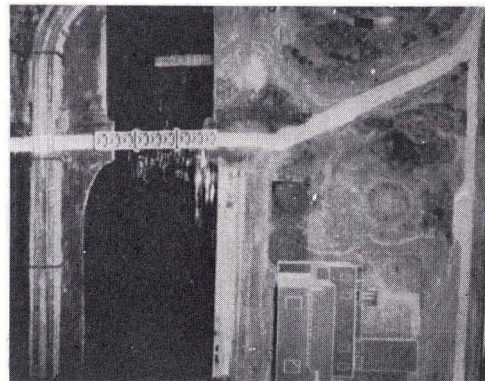
DESIGN OF THE EXPERIMENT

The design of the experiment was guided by a number of practical considerations. First, there were limits as to the total number of subjects available. Then, it did not seem recommendable to study each subject for longer than about three hours. (A preliminary study revealed strong order effects, in that PI's did tend to work faster and faster and, at the end of a long series of tasks, did tend to dismiss a question in few seconds, whereas a similar question would at the beginning of the series have occupied them for some time). On the other hand, our study requires a number of conditions: at least two for stereo (stereo presentation and non-stereo presentation), and at least some for degradation. A final compromise are the three levels of degradation given in the foregoing section, thus six conditions in all. A somewhat reliable estimate of performance under each condition makes it necessary to ask quite a number of questions (especially so since the data are separated into the 3-D categories), so that one hour per subject per condition looked reasonable. This, then, limits the number of conditions to be given to one subject to three (because the total working time was set at about three hours).

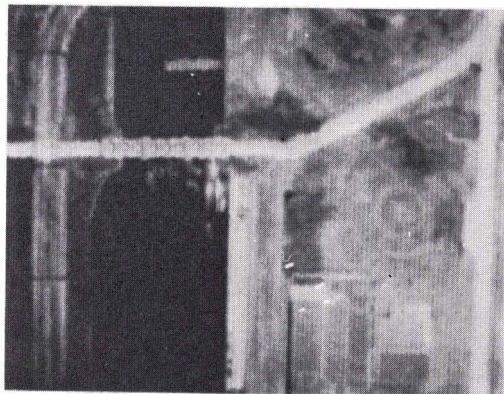
It will be evident that for each condition a new set of questions had to be given to a subject; he cannot be given the same questions over and over again. Therefore we needed



C1



C2



C3

FIG. 5. Example: C1 original; C2 first degradation; C3 second degradation.

three sets of questions, which are called the *A*, *B* and *C* sets, respectively. This division introduces inevitably another independent condition, *sets of questions*, whereas the order in which the sets are given to a subject, also brings up a new condition, technically speaking.

The design finally adopted was a balanced incomplete block design. (Actually, we used plan 11.4 from Cochran & Cox (1957), with the difference that each block was repeated three times in the form of a Latin-Graeco square with order of presentation and sets of questions as additional variables). This design requires 30 subjects. These subjects were trained PI's with extensive experience in stereo as well as non-stereo photo interpretation.

THE QUESTIONS

After some preliminary experiments, three final lists of questions were chosen. Each set was made up of 19 questions, consisting of 4 for detection, 9 for discrimination, and 6 for depth.

PROCEDURE

Each subject received a box with the complete set of 96 photographs in separate envelopes. A Zeiss Aero-type pocket stereoscope was provided, and also an overlay grid. This grid had a coordinate system etched on it which served to specify locations on photographs (in question and in answers). The PI's worked in groups of ten, seated at separate tables. They were provided with a low-watt movable table lamp. Instructions were given in written form; they included a detailed explanation of the whole proceeding with elaborated examples. During the experimental session time recordings were made of the time spent by each PI on each category of questions.

RESULTS

SCORING

In order to achieve an approximately equal contribution of each question to the final score, answers were scored according to a procedure which essentially comes down to dividing the answers in about equal halves of *correct* answers and *false* answers. After this correction, scores were added for each category of questions, resulting in nine scores per subject three categories of tasks times three conditions. Thereafter, a total score was derived from the three category scores by adding the latter after another correction for unequal variances. The whole procedure has the effect that final scores cannot be interpreted in terms of *number of correct answers* or similar realistic notions. Scores can only be compared among themselves, and have no absolute meaning. This is no disadvantage as compared to a score like *percentage of correct*

answers, as the latter score depends largely on the level of difficulty of the questions, which is by itself an arbitrary choice.

An easy and intelligible way of looking at our data is to consider them as relative to the total score which would have been obtained if all answers were correct. Even then, the arbitrariness of level of difficulty remains.

RESULTS

The results are given in Figures 6 to 10 for total scores, scores for detection, discrimination and depth tasks, and total time, respectively. The data were treated by an analysis of variance for incompletely balanced blocks (Cochran and Cox, 1950). This was done for total time, and for accuracy of detection, discrimination and depth tasks separately. The results given in the Figures 6 to 10 are corrected treatment scores, i.e., corrected in the usual way for balanced incomplete block designs. These corrections are very small in most of the cases. The following gives a quick summary of statistically significant results.

Effect of 'sets of questions'. No difference in total time was spent on the tasks, but there are considerable differences for discrimination tasks and also differences for detection tasks. There is no reason for comments upon this result; any difference is only a consequence of the experimenter's arbitrary choice of questions and has no bearing upon PI performance as such.

Order has a very considerable relation with total time spent on the tasks; most of the time is spent on the first tasks. It is, of course, not clear whether this is a matter of learning and becoming familiar with the type of task, or a matter of speeding up because of boredom or loss of interest or both, or any other guess one wants to make. Order also affects accuracy in detection tasks, in the sense of an improvement (the second and third *trials* show im-

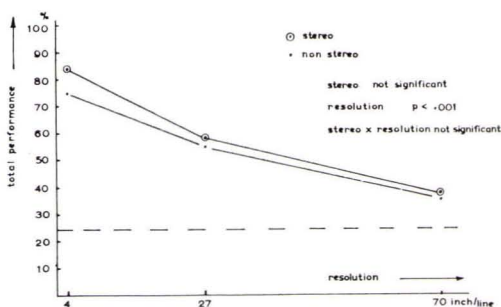


FIG. 6. Total performance accuracy as a function of experimental conditions.

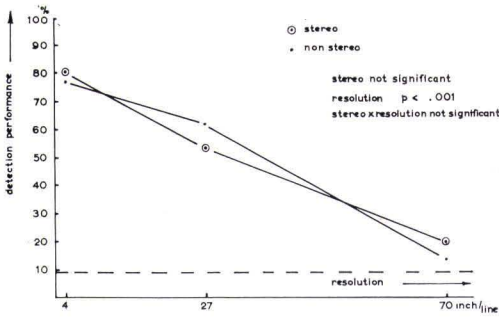


FIG. 7. Accuracy scores for detection tasks.

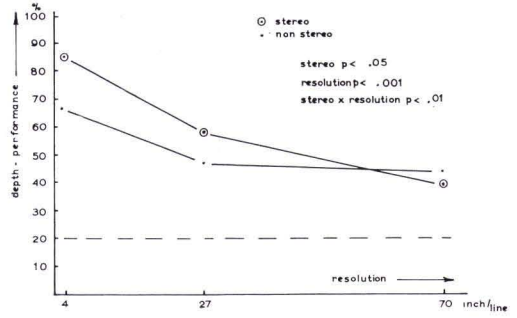


FIG. 9. Accuracy scores for depth tasks.

provements of 23 and 35 percent respectively in relation to the first), but has no effect for discrimination or for depth tasks.

Subjects. There are significant differences between subjects as to total time spent on the task, but not as to accuracy scores.

Conditions are found to have significant effects everywhere. As to total time, both stereo and non-stereo conditions and various levels of degradation are effective. Stereo interpretation takes significantly more time (.001 level of significance), and the other effect is that *undegraded* material takes less time for interpretation (.01 level). There is no difference in this respect between 27-inch and 70-inch levels of degradation.

As to accuracy scores: there is no difference between stereo and non-stereo presentation, except for depth tasks, where the difference attains 5 percent level of significance (Figures 7, 8 and 9). However, there is also an interaction between degradation and stereo in these depth tasks (.01 level) from which can be concluded that this effect of stereo interpretation is only present with undegraded material and disappears completely with the 70-inch degradation. Degradation has a clear-cut effect for all questions (.001 level).

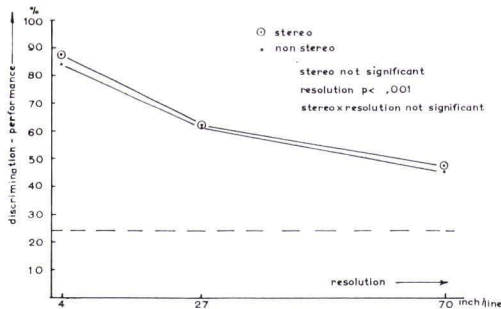


FIG. 8. Accuracy scores for discrimination tasks.

DISCUSSION

RESOLUTION

Resolution appears the important factor in interpretability of photographs. It is seen in Figure 6, that performance drops from a 80 percent level for the undegraded material, to much lower levels for resolutions 27 inches and 70 inches. It is difficult to say to what extent the responses at the lowest level (70 in.) reflect mere guessing from the part of the subjects. In other words, what could a PI have answered our questions if we had not given them any photographs at all? This can be estimated only very roughly. The estimation of chance level in detection tasks is the most reliable one, and if we take these tasks as a norm, it is seen that performance deteriorates from 80 percent at 4-in. photos to about 15 percent at the 70-in. photos.

Least affected by degradation are *orientation* questions (e.g., the direction of movement of a tank); it seems that minimal cues are needed for discerning the orientation of familiar objects. It is worth noting that the

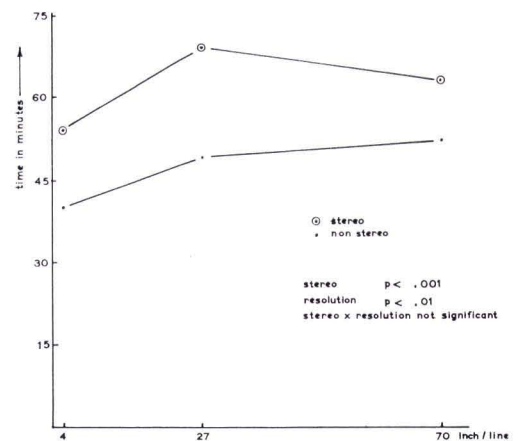


FIG. 10. Total time scores averaged over subjects.

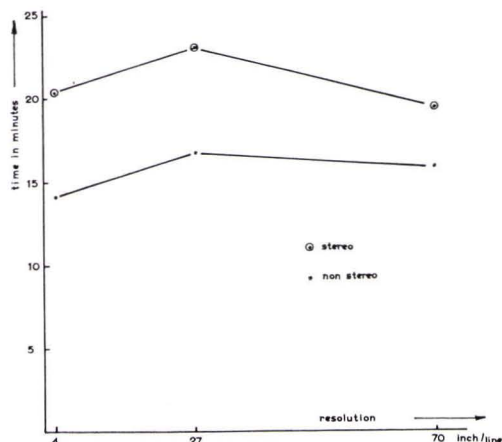


FIG. 11. Time scores, averaged over subjects, for depth tasks only.

direction of scanning lines seems to have no effect here as the objects to which the questions refer vary in this respect. We may add that in the preliminary experiment in which the questions were tested, we found that orientation questions tended to be too easy; even at high level of degradation no failures occurred.

Finally, we may comment on the relationship between degradation and interpretation time. Figure 10 shows that there is an increase of interpretation time if one goes from 4-in. photos to 27-in. photos. There is no further increase in time if one compares 27-in and 70-in. photos (the line-structure is absent in the untreated photographs of 4-in. resolution). However, it is only natural to assume that interpretation time will drastically decrease if degradation is continued; at the level where there is literally nothing to see, no PI will spend much time on this.

It is not too daring therefore, to assume that interpretation time and degradation level have a curvilinear relation in that there is a maximum time at some level of resolution. The argument is supported by looking at depth tasks only. Figure 9 shows that stereo information is only effective for these tasks at levels 4 in. and 27 in.; at the 70-in. level it is absent. Therefore, one should expect that PI's do not spend much time in trying to pick up stereo information at the 70-in. level, so that at this level their total time score for depth tasks would be lower than that for the 70-in. level. This is corroborated by Figure 11, where the stereo curve indeed shows a maximum at the 27-in. level, whereas this trend is

much less manifest in the non-stereo presentations.

STEREO

Our second main question was whether stereo presentation of photographs is useful for the PI. The results of our experiment show that stereo presentation has some advantage for depth tasks. If one compares Figures 7, 8 and 9, it will be noticed that stereo has no significant effect for detection or discrimination tasks, whereas for depth tasks there is a significant interaction which can be interpreted to mean that stereo presentation gives better results at the 4-in. level and at the 27-in. level; at resolution level of 70 in. the advantage of stereo has disappeared. The result means that stereo presentation may have its advantages for the height estimation of objects with a specific height, and only at moderate degrees of degradation. For detection and discrimination tasks, stereo presentation does not seem to be useful at all.

A decision on the use of stereo presentation should also take account of the working speed of PI's. If we look at the time scores in our experiment it is seen that stereo interpretation takes much more time than non-stereo work. The difference is significant not only with depth tasks (where the increase in time is at least rewarded by some increase in accuracy), but also with detection and discrimination tasks, where it is doubtful whether stereo presentation has any use. At the average, interpretation of stereo material takes 32 percent more time than interpretation of single photographs (in a preliminary study an increase of 45 percent was found).

In our opinion, this is a decisive argument against the use of stereo pictures for object identification. Even if stereo presentation would result into a somewhat better accuracy, the fact that so much more time has to be allowed makes stereo presentation unprofitable. Only if there are no practical limits to the PI's time, stereo might bring out its use (if there is any) but it seems most unrealistic to assume that PI's abound in time, particularly now that modern reconnaissance techniques tend to bring up so much material for interpretation at such a high speed that the real problem seems to be how to digest this abundance.

The only possible flaw in our argument might be that there is an intricate relation between accuracy and time spent on a photograph. Namely, (and it is not overlooked) that the advantage of stereo might show up again if interpretation time were restricted,

and if PI's were instructed to work as quickly as possible in the restricted time allowed.

CONCLUSIONS

1. PI performance deteriorates with degradation of the photograph. The amount of deterioration depends upon the type of target, the type of task and the number of lines passing the target.

2. Interpretation time increases with 25 percent from 4-in. resolution level to a 27-in. level; no increase was found from a 27-in. to 70-in. level. It is assumed that interpretation time has a maximum at some level of resolution. At first, time tends to increase with diminishing resolution, but when further degradation deletes the information content, time decreases again.

3. Stereo presentation seems useful only for height estimation of objects without specific height, and this only for moderate degrees of degradation. Current military PI-work does not often contain the type of questions where stereo presentation is useful.

4. The pertinent disadvantage of stereo presentation is that it does increase time spent on photographs with 32 percent.

REFERENCES

- Cochran, W. G. and Cox, G. M., 1950, *Experimental Designs*, Wiley & Sons, New York, 611 pp.
- Elias, P., Grey, D. S. and Robinson, D. Z., 1952, "Fourier Treatment of Optical Processes", *J. Opt. Soc. Amer.*, 32, p. 127.
- Fry, G. A. and Enoch, J. M., 1958, "Human Aspects of Photographic Interpretation," RADC TN-58-298, MCRL-7, *I.T.R.* nr. (696)-10-266, A.F. 30-(602)-1580.
- Hempenius, S. A., 1964, "Aspects of Photographic Systems Engineering," *Appl. Opt.*, 3, p. 45.
- Langner, G. and Müller, R. 1963, "Zur Messung der Schärfe photographischen Aufnahmematerials mit physikalischen Methoden," *Phot. Korr.*, 99, p. 35 and 69.
- Radio Corporation of America, 1958, "An Exploratory Image-Interpretation Experiment," *RCA* 2677 (pr. 1, R.Cy.M.-W-2).
- Sadacca, R., 1963, "Human Factors in Image Interpretation," *Phot. Eng.*, 24, p. 978.
- Schade, O. H. "Image Gradation, Graininess and Sharpness in Television and Motion Picture Systems," *Journal SMPTE*, 1951, 65, p. 137; 1952, 58, p. 181; 1953, 61, p. 97; 1955, 64, p. 593.
- Zeidner, J. and Birnbaum, A. H., 1959, "Survey of Psychological Factors in Image Interpretation," PRB Res. St. 59-1, US Army Prog. Number 2L 95-60-001.
- Zeidner, J. 1961, "Requirements for Research on Psychological Factors in Image Interpretation," in *Vision Problems of the Armed Forces*, Whitcomb, M.A., ed., AF-NRC Committee on Vision p. 158.
- Zeidner, J., Sadacca, R. and Schwartz, A. I., 1961, "Human Studies in Image Interpretation: The Value of Stereoscopic Viewing," HFRB, TR note 114, DA Project 2L 95-60-001.

Notice to Authors

1. Manuscripts should be typed, double-spaced on $8\frac{1}{2} \times 11$ or $8 \times 10\frac{1}{2}$ white bond, on *one* side only. References, footnotes, captions—everything should be double-spaced. Margins should be $1\frac{1}{2}$ inches.
2. *Two* copies (the original and first carbon) of the complete manuscript and two sets of illustrations should be submitted. The second set of illustrations need not be prime quality.
3. Each article should include an abstract, which is a *digest* of the article. An abstract should be 100 to 150 words in length.
4. Tables should be designed to fit into a width no more than five inches.
5. Illustrations should not be more than twice the final print size: *glossy* prints of photos should be submitted. Lettering should be neat, and designed for the reduction anticipated. Please include a separate list of captions.
6. Formulas should be expressed as simply as possible, keeping in mind the difficulties and limitations encountered in setting type.