Automatic Photo Reading

J. K. HAWKINS and C. J. MUNSEY,
Research Lab., Aeronutronic Div., Ford Motor Co.,
Newport Beach, Calif.

ABSTRACT: *Because of the large volume of raw data possible of collection by means of aerial photography, the automatic detection and counting of simple objects or terrain features from the raw image is an attractive possibility for relieving some of the routine burden on photo interpreters. This paper describes one economical automatic technique for performing such routine tasks. It also presents the details of illustrative examples.*

Previous approaches to this problem have involved either the use of some form of optical correlation process, or some form of complex electronic "brain" analog. The former in general suffers from the lack of sufficient discrimination ability, while the latter generally suffers from a serious loss of resolution. Optics, combined with the use of photographic film in a multistage, nonlinear decision process, offers advantages over both of these approaches. The image resolution attainable with photography is largely retained, while the complex logical decisions possible with electronics are readily performed.

AUTOMATIC IMAGE PROCESSING

A LARGE number of diverse approaches have been taken to the problem of automatically detecting or recognizing certain objects or features in image data. Frequently the goal is to detect, locate and count well-defined objects such as specific classes of military targets which may appear on voluminous aerial reconnaissance photography. Another objective is to detect and make gross measurements of the amount of some terrain feature such as wooded areas or urban areas on each of a large set of aerial photographs. In both cases, the raw data are customarily in photographic form.

At one extreme of the types of processes applied to such data are the optical or electronic correlation techniques, which have been investigated by a number of authors.[1,2,3] These techniques generally involve the matching of the unknown image against some standard. They can be characterized as follows: 1) a linear process (one which obeys the laws of superposition; the correlation integral) is applied, and 2) a single nonlinear, yes-no decision process is performed upon the results of (1). The fact that only a single decision is made inherently limits the sophistication of this general type of approach.

At the other extreme lie a great variety of complex electronic network techniques which have been described by several authors.[4,5,6] Such techniques can, in principle at least, carry out the sequences of decision processes necessary for sophisticated recognition. The use of conventional electronic circuits, however, inherently limits the number of image elements which can be treated simultaneously, and therefore limits either the resolution or speed which can be attained.

In order to take advantage of the best features of both the above approaches, what is needed is a method for performing very large numbers of yes-no decision processes—the number corresponding to the resolution elements in image data—at reasonable cost. One fulfilment of this need is photography itself. The use of photography to perform what amounts to computer calculations has been described by at least two authors.[7,8] The term eulogismography was apparently coined by Million to describe the calculating function of the photographic process.

What has been overlooked, however, is the importance of the very thing which is frequently one of the drawbacks of photographic film, namely, its nonlinearity. In very high-contrast films in particular, this nonlinearity amounts to a yes-no decision process. That is, photographic film may be regarded as asking, at every resolution element in the film: is the input (light) greater than or less than some standard? If it is greater, the resulting negative transmittance in that area of the film approaches zero. If it is less, the resulting negative transmittance approaches one. In elec-

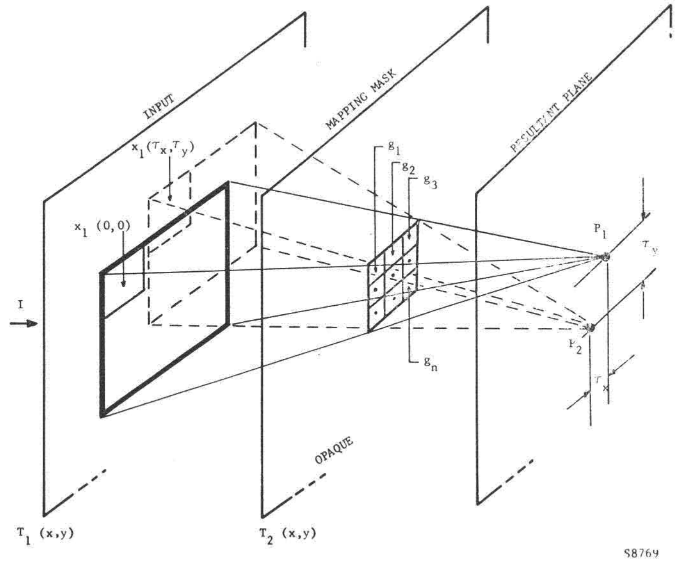

FIG. 1. Image-wise local decision process.

tronic circuits this is known as a binary or two-state (0 or 1) decision.

It will be shown how sequences of such decision, performed by photographic films upon image data, can accomplish the detection of certain shapes. The basic philosophy behind the processing sequences is that of building up a "logical" chain of events, similar in some sense to a rudimentary "reasoning" sequence. It will be seen, therefore, that at each step of the process, the photographic system is in effect making a large number of decisions of the form, "If . . . then. . . . " While the examples presented will be simple, the extension of this technique to lengthier sequences required for more sophisticated detection problems will be readily appreciated by the reader.

Before describing processing details, it is well to point out that photo interpreters have long recognized the importance of logical or decision processes. This is implicit in the existing carefully worded and documented descriptions of clues for recognizing a great variety of objects, features and events. It is also implicit in the quantitative approach by Ray and Fischer[9] to geologic interpretation. This close correspondence between logical processes and photographic interpretation techniques was first brought to the authors' attention by J. E. Gillis and B. B. Scheps of GIMRADA, U. S. Army Corps of Engineers, Ft. Belvoir, Virginia. It was their viewpoint which gave impetus to the studies leading to the results described below.

PRINCIPLES OF OPERATION

The type of logical decision which it is appropriate to make in automatic processing of image data is generally characterized by two properties: 1) each decision in a sequence is as simple in nature as possible, and 2) each decision depends only upon the density pattern in a local area of the image. The ultimate decision to be made, of course, is of the type: "yes or no: there is an object (e.g. target) of the specified type is located here (a point or local region on the image)." This is much too complicated a decision to perform in a single automatic processing step. Rather, it is necessary to build up to this final decision by means of a series of simple sub-decisions.

For automatic processing, each sub-decision must be capable of being performed in a single processing step. This restricts the nature of sub-decisions which it is possible to make. Using any type of amplifying device, including high-contrast photographic film, such sub-decisions can only be of the form: "Is the sum of the image transmittance, weighted at each point by some appropriate factor and the sum taken over a local region of specified shape, greater or less than a given amount?" This question must be asked at each point of the image.

An optical-photographic system which mechanizes this single-stage decision process is illustrated in Figure 1. The input transparency (which may be either the original image, or the result of some previous decision

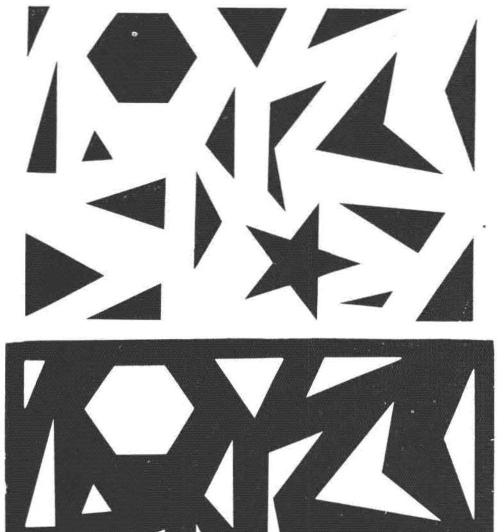

FIG. 2. Field of geometrical shapes.

process) is T_1. It is illuminated from behind by a uniform diffuse light source. The local weighting function is performed by transparency, T_2, which is opaque everywhere except within a region corresponding to the specified shape. Within that region, it possesses various transmittance values, $T_2\ (x,y)$ such that the specified weighting function is performed. A high-contrast film is placed in the position of the resultant field.

It can be seen that at every point on the resultant field the amount of light energy is, assuming discrete resolution elements in T_1 and T_2,

$$I = \sum_{i=1}^{N} g_i x_i \qquad (1)$$

in which

g_i = transmittance of ith resolution element of T_2, and

x_i = transmittance of ith relative resolution element of T_1.

If the transparencies are regarded as continuous rather than discrete, then (1) becomes

$$I(\alpha, \beta) = \int T_2(x, y) T_1(x + \alpha, y + \beta) dxdy. \qquad (2)$$

Equation (2) is of the form of the convolution or correlation integral. Its extent (the limits of non-zero integration) however, is confined to the local area defined by the non-zero transmittance of the shape-and-weight-defining transparency, T_2.

It can be seen that images conforming to equations (1) or (2) are formed simultaneously all over the field with respect to every local region. Thus at P_1 the weighted sum of input transmittances indicated by the solid square on T_1 is formed, while at P_2 a similar sum is formed with respect to the dotted square on T_1. Therefore, a developed high-contrast film placed on the resultant field simultaneously answers the series of questions, "Is the transmittance (weighted and shaped by T_2) around a point on the input image corresponding to P_1 more or less than θ (a given amount); is the transmittance around a point on the image corresponding to P_2 more or less than θ; . . . is the . . . corresponding to P_i more or less than θ?" The index, i, corresponds to all of the resolution elements in the input image.

The power of this process resides in two factors: 1) the huge number of decisions (of a similar nature) which are made in a single exposure, and 2) the fact that the output of one processing step (the high-contrast film after development) can serve as the input (T_1) to a subsequent step enabling long sequences of decisions to be carried out.

More sophisticated optical systems exist for performing the above process; however, they do not affect the operations (1) or (2) which are performed. It is also worth noting that the same film (resultant field) may be exposed to more than one T_1–T_2 pair. This is particularly valuable since it can be shown that most useful sub-decisions require both an input image and its negative. This corresponds to requiring negative values of g_i in (1) or negative T_2 in (2). Since negative transmittance is not available except in coherent optical systems, the same effect can be obtained by using the negative (reversal) of the image.

In the following illustrations, decision functions (T_2) will be specified by a drawing which defines the shape of the non-opaque region. Within this region, signs will indicate whether the positive (+) or negative (−) of the given input transparency is to be used.

ILLUSTRATIVE DECISION SEQUENCES

Since geometrical shapes are of interest in object detection, a simple test was conducted on the artificial field of arbitrary shapes shown in Figure 2. Object of the test was to detect acute angles, whether they be white on black or black on white. Some investigation showed that the decision function of Figure 3(a) would be adequate. It can be seen that the amount of light passing through this pair of transmittance shapes is a function of the angle of the object appearing in alignment with the function, as shown in Figure 3(b) and (c). The form of this quantity of light, as a function of angle, is shown in Figure 4. Clearly, a decision made at a level corresponding to 90° and another at a level corresponding to 270°, when combined, will produce the desired overall decision. Note that either less than 90°, *or* greater than 270°, meets the criterion for acute black-or-white angles. The complete sequence of decision processes is therefore as shown in Figure 5. An example of the results of the angle decision process, in one polarity (black-white direction) is shown in Figure 6.

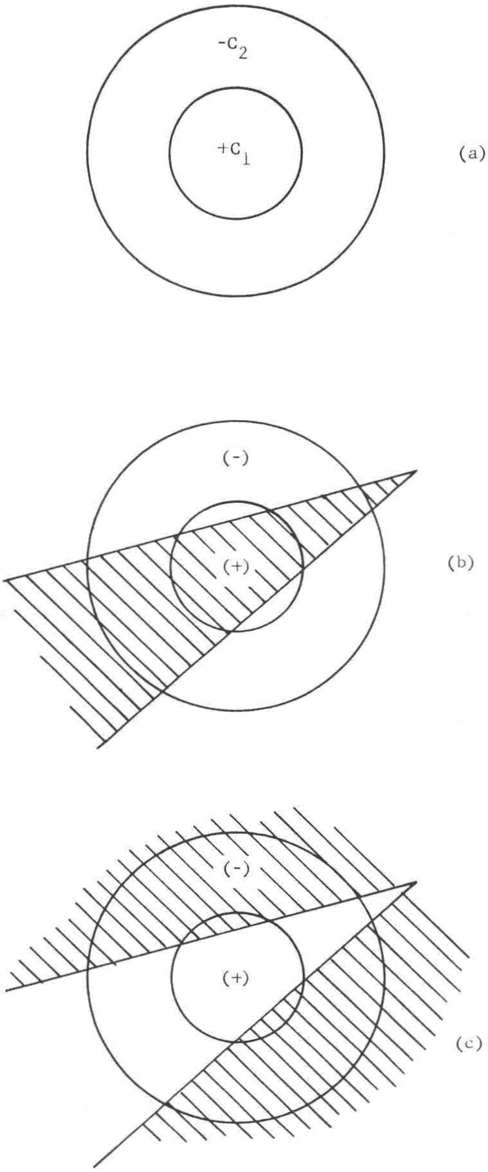

FIG. 3. Decision function for angle detection. (a) Specification of decision function. (b) Angle less than 90° on decision function. (c) Angle greater than 270° on decision function.

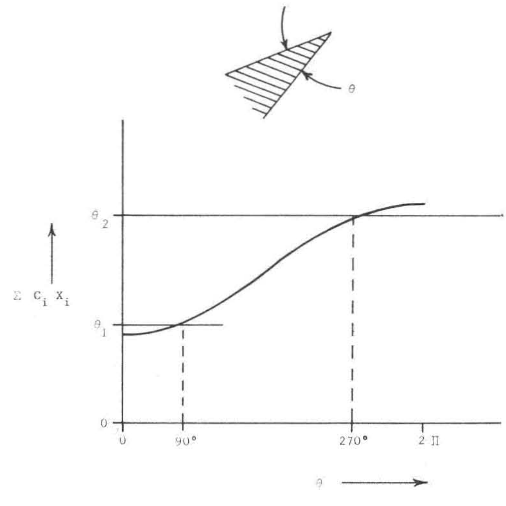

FIG. 4. Output of decision function versus local angularity.

A "yes" answer in a decision process of this type simply corresponds to a small dot located at the appropriate point on the film. Since dots are hard to see, a final step in the process consisted of generating small circles (or any other arbitrary indicator) surrounding "yes" decisions. This was done by placing a donut shaped transparency at T_2, and the dot-result transparency at T_1. The final output is shown in Figure 7 and the superposition of the input and result in Figure 8. Note that some errors have been made, notably of omission. No detailed analysis has been performed, but it is believed that these are due to the rather crude optical setup employed, to dust, glass imperfections and inadequately controlled development. An interesting error occurs in approximately the center of each half-field, where two triangles merge into something like a thin line locally. Some consideration of Figure 3 will show that it is inadequate to distinguish between thin lines and acute angles; a more sophisticated decision process would be required to make this distinction.

It is worth noting that the function-pair of Figure 3 is similar to the "unsharp mask" technique long known to photographers, or to the photoline process[10] used to enhance

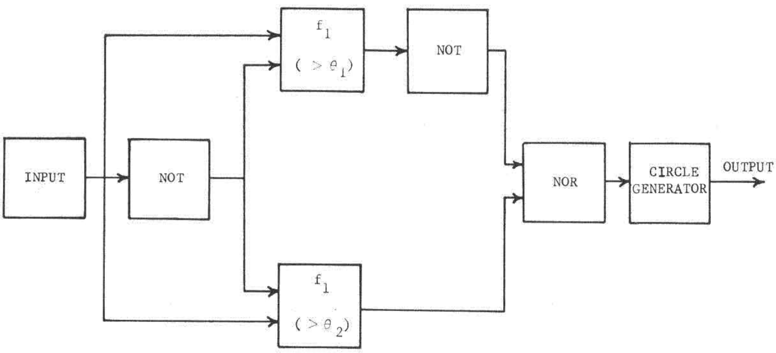

FIG. 5. Decision sequence for acute angle detection process.

contrast. It may also be looked upon as a first approximation to two-dimensional differentiation, or to the impulse-response function (sine x/x).

In real image data, inputs seldom possess the ideal black-and-white binary values of Figure 2, and the background is seldom so uncluttered. Therefore experiments were conducted on a portion of a typical aerial photograph, shown in Figure 9. The object of the experiment was to find a processing sequence which would detect vehicles of standard passenger-car size. After some study of vehicle images and initial experimentation, it was concluded that a major fraction (more than 85%) of the vehicles in this photograph could be detected by a straightforward process of testing (by decision processes) for appropriate line-pairs of various polarities (black-white reversals) in about eight discrete ranges of angular direction. A complete sequence of decision processes would, it is believed, be able to take into account two-tone vehicles of both polarities as well as solid-color vehicles of both polarities. A complete sequence is rather tedious, however, so in order to perform a reasonable test only a portion was actually carried out. This truncated version of the decision sequence is described below.

Because of a certain overlap between functions performed at 90° relative rotation, the simplest possible process consisted of detecting vehicles aimed in mutually perpendicular directions. Because of the predominance of

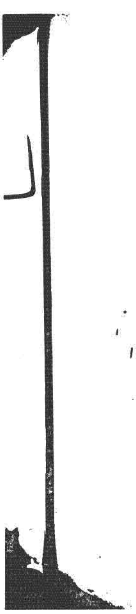

FIG. 6. Result of one stage in decision sequence.

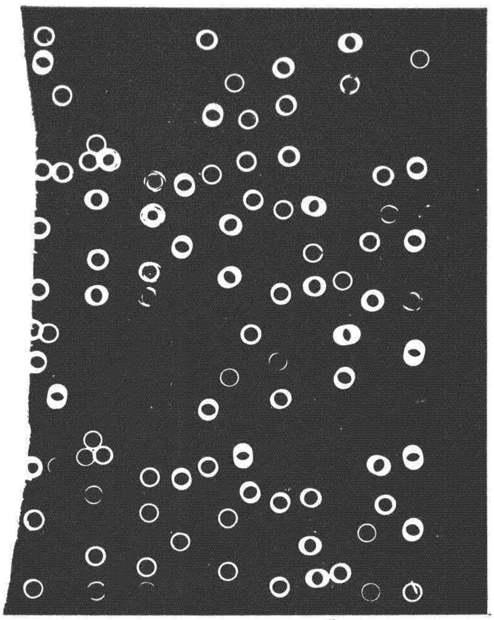

FIG. 7. Final output of decision sequence.

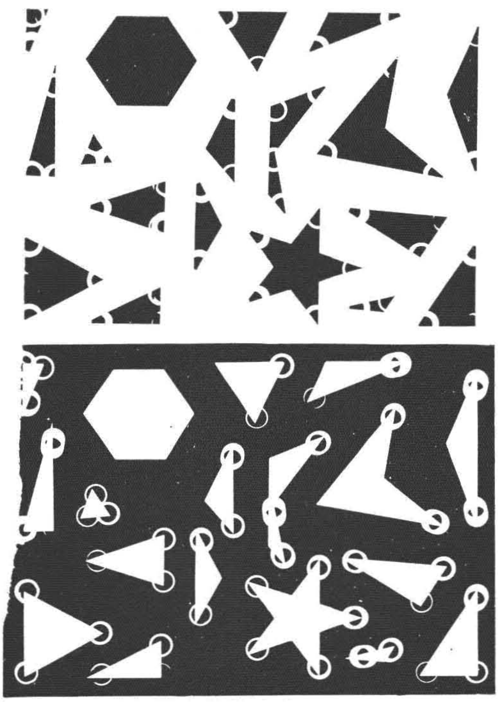

FIG. 8. Superposition of input field and final output.

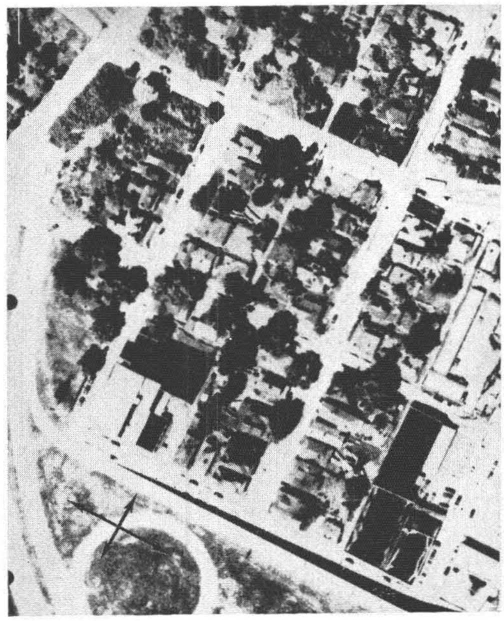

FIG. 9. Aerial photograph input field.

roads, these directions were chosen in accordance with the arrows on Figure 9. Attention was also restricted to dark vehicles. Processing may be listed in a verbal description as follows:

(1) *Long edges (corresponding to the length of vehicles) were detected* by means of the transparency-pair function, $f_1(\theta)$, of Figure 10(a). An example of this decision applied to Figure 9 is shown in Figure 11 for $\theta=0°$ (see arrows).

(2) *Pairs of (1) of the proper polarity for dark vehicles, and separated by approximately the width of a vehicle were detected.* That is, only if $f_1(\theta)$ and $f_1(\theta+180°)$ are "yes" at an appropriate separation, according to the function-pair of Figure 10(b), is the decision "yes." Call this decision $f_3(\theta)$. An example of the output of this decision step for $\theta=0°$ is shown in Figure 12.

(3) *Short edges (corresponding to the width of vehicles) were detected* by means of a transparency pair similar to f_1, but of different dimensions, as shown in Figure 10(c). Call these $f_2(\theta)$. An example of this decision applied to Figure 9 is shown in Figure 13 for $\theta=90°$.

(4) *Pairs of (3) of proper polarity and separated by approximately the length of a vehicle were detected.* That is, only if $f_2(\theta+90°)$ and $f_2(\theta+270°)$ are "yes" at an appropriate separation, according to the function-pair of Figure 10(d), is the decision "yes." Call this decision $f_4(\theta)$. An example of the output of this decision step for $\theta=90°$ is shown in Figure 14.

(5) A combination of (2) and (4) was taken, such that only if $f_3(\theta)$ and $f_4(\theta+90°)$ was present was the decision "yes."

(6) A circle-generator was applied to (5).

Some details of the processing inherent in film reversals are left out of the above list for simplicity. The results of step (6) are shown in Figure 15, and the superposition of Figure 9 and 15 is given in Figure 16. Some distortions due to the crude optical system employed have occurred at the lower left corner, and caused the circle to miss the correctly detected vehicle in the lower right corner. Nevertheless, a significant proportion of the non-two-toned vehicles aimed in the two given directions has indeed been picked out by the process. In addition, a vast amount of background clutter corresponding to houses, trees, backyards, etc., has been correctly rejected.

It is worth noting that while the above does not represent the only possible solution to the stated problem, some simpler approaches have been tried and found to fail. For

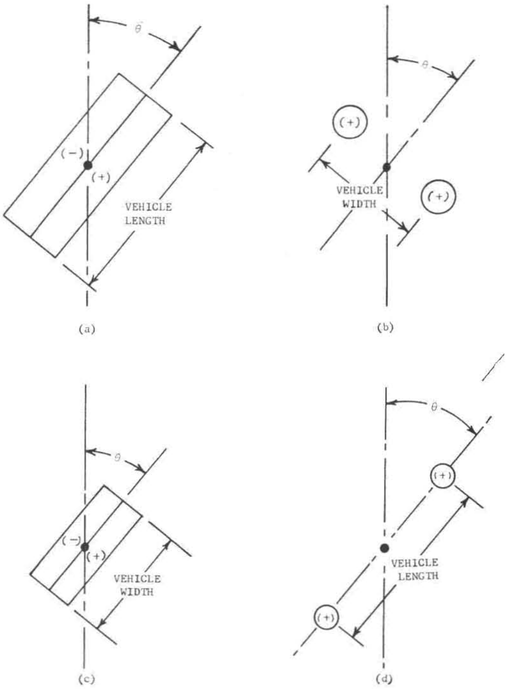

FIG. 10. Decision functions for vehicle detection sequence. (a) Long edge detector. (b) Close parallel edge-pair detector. (c) Short edge detector. (d) Far parallel edge-pair detector.

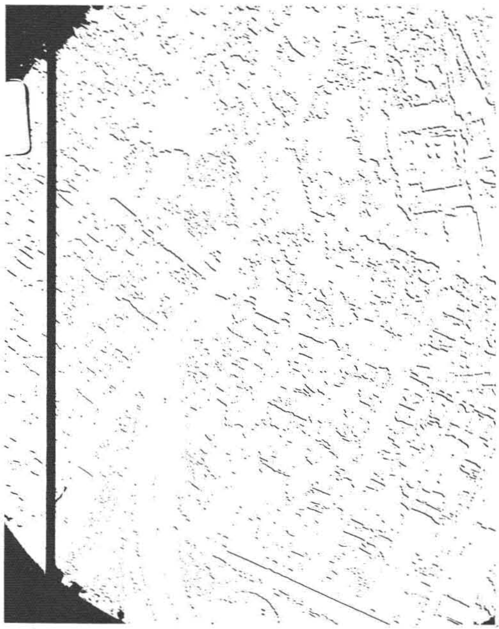

FIG. 11. Result of long edge stage in decision sequence.

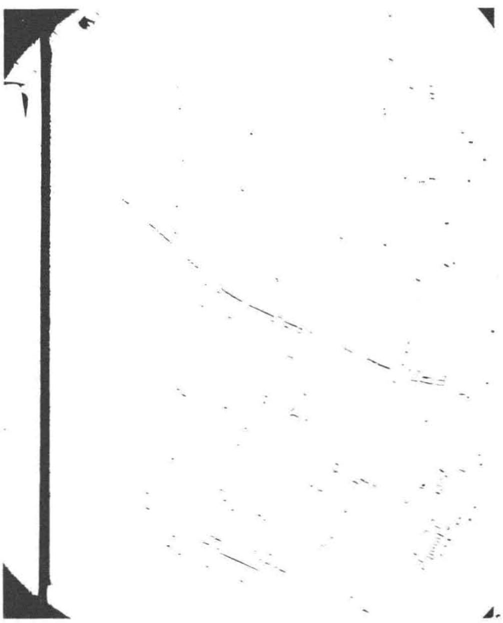

FIG. 12. Result of close parallel long edge-pair stage in decision sequence.

FIG. 13. Result of short edge decision stage.

FIG. 14. Result of far parallel short edge-pair decision stage.

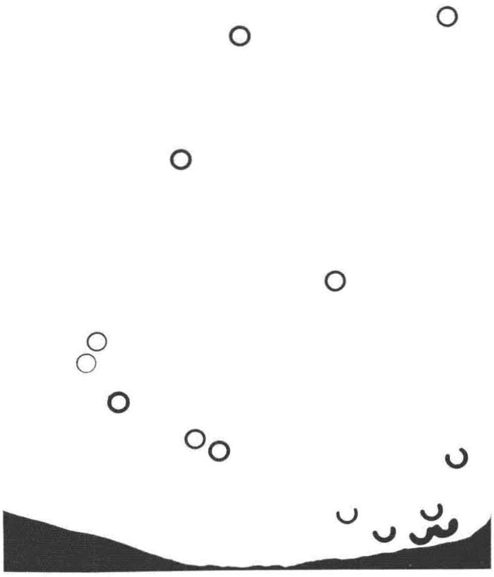

FIG. 15. Final output of vehicle detection decision sequence.

FIG. 16. Superposition of input aerial photograph and final output.

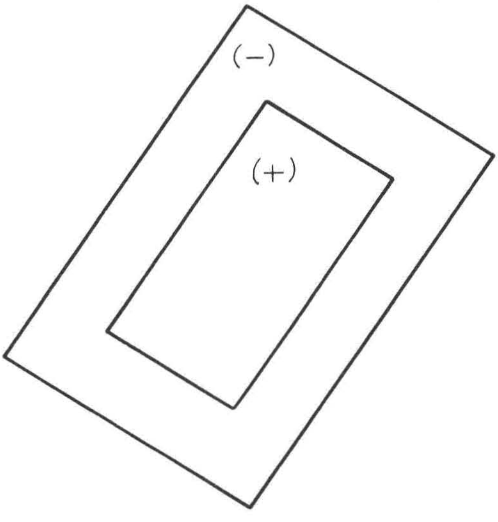

FIG. 17. Single-stage ideal oblong-detector decision function.

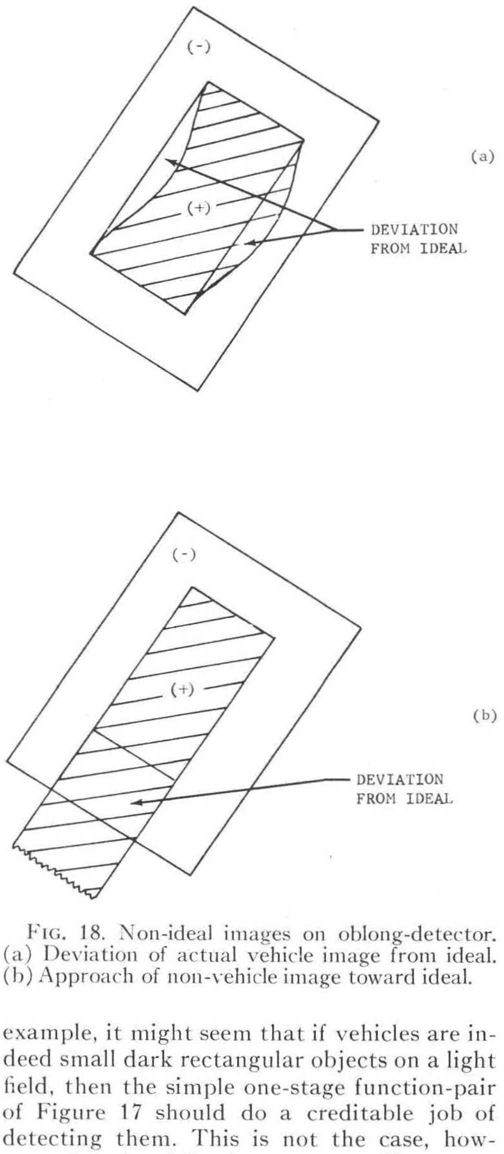

FIG. 18. Non-ideal images on oblong-detector. (a) Deviation of actual vehicle image from ideal. (b) Approach of non-vehicle image toward ideal.

example, it might seem that if vehicles are indeed small dark rectangular objects on a light field, then the simple one-stage function-pair of Figure 17 should do a creditable job of detecting them. This is not the case, however, for the following reasons. The worst vehicle image fails to "match" the ideal of Figure 17 by some percentage. Therefore, the decision level must be set somewhere below the total relative area of Figure 17. An example is given in Figure 18(a). On the other hand, some objects in the photograph come close to matching the ideal rectangular vehicle image. Such an object is the dark shadow of approximately the width of a vehicle appearing about half way between dead center and the lower righthand corner of the picture, Figure 9. It possesses nearly square corners, and deviates from a rectangular ideal only in that it extends too far, as shown in Figure 18(b). If its deviation from the ideal is no greater than that of Figure 18(a), the two cannot be distinguished in a single processing step. It appears that aspect-ratio discrimination, which comes so naturally to human beings, requires more than one decision stage in an image processing sequence of the above type.

REFERENCES

1. Cutrona, L. J., *et al.*, "Optical Data Processing and Filtering Systems," *IRE Trans. on Information Theory*, Vol. IT-6, pp. 386–400; June, 1960.
2. Roetling, P. G. and H. B. Hammill, "Study of Spatial Filtering by Optical Diffraction for Pattern Recognition," Report VE-1522-B-1. Cornell Aeronautical Lab., Buffalo, N. Y., Feb., 1962.
3. "Final Report: Electronic Correlation Techniques for Target Recognition," Report GER-10449. Goodyear Aircraft Corp., Akron, Ohio, Jan., 1962.
4. Bledsoe, W. W., and I. Browning. "Pattern Recognition and Reading by Machine," Proc. EJCC, pp. 225–232; 1959.
5. Chow, C. K., "An Optimum Character Recognition System using Decision Functions." *IRE Trans on Electronic Computers*, Vol. EC-6, pp. 247–254. Dec., 1957
6. Murray, A. E., "Perceptron Applications in Photo Interpretation," PHOTOGRAMMETRIC ENGINEERING, Vol. XXVII, No. 4, pp. 627–637, Sept., 1961.
7. Howell, B. J., "Optical Analog Computers," *Jour. of the Optical Soc. of America*, Vol. 49, 10, pp. 1012–1021, Oct., 1959.
8. Million, C., "Photographic Methods of Calculation (Eulogismography)", C. R. Acad. Sci., (Paris), Vol. 234, pp. 313–316, Jan. 14, 1952.
9. Ray, R. G. and Wm. A. Fischer, "Quantitative Photography—a Geologic Research Tool," PHOTOGRAMMETRIC ENGINEERING, Vol. XXVI, No. 1, pp. 143–150, March, 1960.
10. Clair, J. St., and M. Merriam, "The Use of Photography to Achieve Basic Principles of Map Design," PHOTOGRAMMETRIC ENGINEERING, Vol. XXVI, No. 3, pp. 498–504, June, 1960.