Journal of the ACMVolume 9Issue 2April 1962 pp 259–267

Operations Useful for Similarity-Invariant Pattern Recognition

W. DOYLE*

The RAND Corporation, Santa Monica, California

Summary. Some ideas for position- and size-invariant two-dimensional pattern recognition are discussed. They lead, in some cases, to easily mechanized operations. An application to detection of straight lines is proposed.

I. *Introduction*

Many schemes for recognition of plane patterns are variations or elaborations on the following: an image field is divided into resolvable spots, most commonly a rectangular array of equal squares. A pattern to be considered is given as a list of numbers, $I(x, y)$, one number to each spot, representing image transparency or reflectivity at each picture element. For each possible pattern a standard list is available (mask or reference function) and the identification of an unknown is accomplished through its cross correlations with these standards. The authors often note that this method inherently regards as different patterns two images that are translations of one another or that can be made to coincide by a rigid motion. Thus, in order to recognize the same configuration in various positions multiple reference functions (or movements of a basic one) are necessary with this simple scheme. When the number of possible positions is small or when the movements and multiple comparisons can be made easily and quickly, this replication of references is not onerous and forms part of some successful mechanisms. When the permissible field distortions are to include rotation and dilatation in addition to translation, the motions or multiplicity of references may become impractical.

The excessive cost of the procedure just discussed may be ascribed to the fact that it actually gives much more information than is usually desired. Thus, in a case, for example, in which a field is to be examined for the presence of line segments, the straight cross correlation method leads to a separate decision for each possible position of a line in the frame. In many cases, however, and this establishes the viewpoint of this paper, what is sought is an answer to the much vaguer question whether there is any segment somewhere in the field. What might suffice to answer this sort of question is some pattern transformation or processing of the original image field involving less work than multiple cross correlation and yet such that the transformations of all variations of the same

* Any views expressed in this paper are those of the author. They should not be interpreted as reflecting the views of The RAND Corporation or the official opinion or policy of any of its governmental or private research sponsors. Papers are reproduced by The RAND Corporation as a courtesy to members of its staff. Parts of this note summarize work done at Aeronutronic, Newport Beach, California.

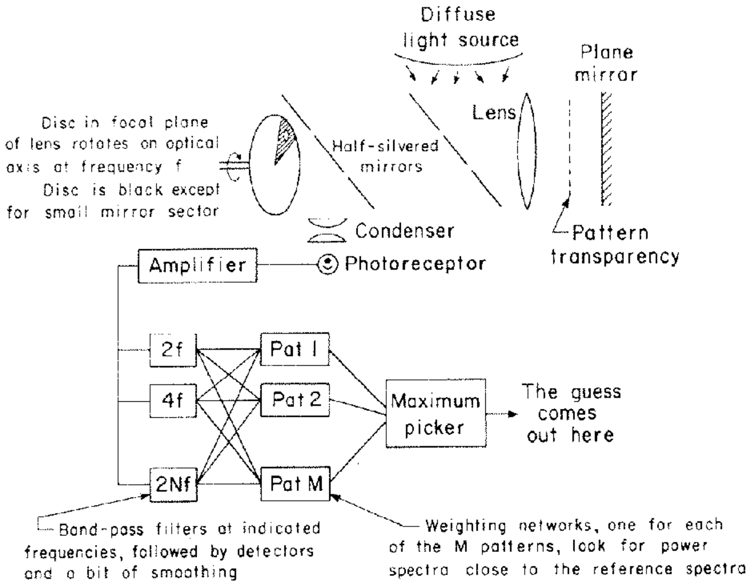


FIG. 1. A similarity-invariant pattern classification device

pattern show up in the same output bin. Such processing is a form of abstraction insofar as a partition of the set of possible input fields into a set of classes is replaced by a partition into fewer or more extensively populated classes. After such a transformation, cross correlation may be a practical way to identify the unknown classes.

II. *Spectra and Filters*

A trivial but interesting example of the sort of processing proposed above is furnished by the division of the electromagnetic spectrum into the narrow bands used in ordinary radio communications. The patterns (one-dimensional functions of time) that turn up in such an output bin (filter) are those whose spectral components lie in a certain frequency interval. An analogous abstraction, the two-dimensional spatial frequency spectrum, has been suggested for use in recognition or detection of plane patterns [1]. Both this and its Fourier transform,[1] the two-dimensional autocorrelation function $C(h, k)$, are independent of translations of the original image plane just as the autocorrelation and spectrum of an electrical signal are independent of time delays of the original wave form. Thus an operation is already available to transform the input image plane (nonlinearly) in such a way as to produce a new plane distribution independent of translation of the original. Furthermore, the autocorrelation is obtainable from a transparent input image by a very simple optical arrangement [2], which is

[1] For a general introduction to these ideas and some interesting applications see also E. O'Neill, "Spatial Filtering in Optics," *Trans. IRE-PGIT*, *IT-2* (June 1956), 56–65.

part of Figure 1. It thus seems likely to form a useful basis for invariant pattern recognition or detection.

The same general principles may also be applied to derive further abstractions that are independent of other transformations of the original image field. We start with the two-dimensional auto-correlation,

$$C(h, k) = \iint I(x, y)I(x - h, y - k)dx\,dy,$$

of the input image $I(x, y)$ and consider the effect on $C(h, k)$ of rotating $I(x, y)$ around any (x_0, y_0). Evidently the effect is merely to rotate the corresponding autocorrelation function $C(h, k)$ about $h = k = 0$. To obtain from $C(h, k)$ a further abstraction that will be independent of rotation of $C(h, k)$ about $(0, 0)$, hence independent of rotation of $I(x, y)$ about any center, the analogous spectrum corresponding to shifts of polar angle in the autocorrelation image plane may be taken. Since $C(h, k)$, expressed in terms of polar coordinates in the (h, k)-plane as $C(r, \theta)$, is periodic in θ (the period is π, in general), the corresponding power spectrum is formed by the squared magnitudes of the Fourier series harmonics. Clowes and Parks [3] have used $C(r_0, \theta)$, for fixed r_0, as a position-invariant abstraction from $I(x, y)$.

So far nothing has been said about r. If the final result of the input data transformation is to be independent of size, r cannot be left as a free parameter. Various courses are possible and three will be described. When this subject was first considered, the variable r was simply averaged out and attention fixed on the function[2] $f(\theta) = \int_0^\infty C(r, \theta)r\,dr$ with its harmonics normalized to unity total power. This was the first course. Crude as it may seem it forms a data reduction mechanism which was quite successful when tested by digital computer simulation. With radial variations of $C(h, k)$ averaged out the scheme is evidently well adapted to recognize patterns consisting of a small number of straight lines and for the test mentioned it was used to classify randomly shrunken, rotated and shifted line segments, triangles and squares in the presence of noise equivalent to slightly under unity signal-to-noise ratio.[3] A mechanization of this scheme is

[2] Roger Stafford of Aeronutronic immediately commented that any other radial moment (or all of them) would also serve.

[3] In the course of the computer experiments described in the Appendix it was noticed that a considerable performance improvement resulted from excluding from further processing those autocorrelation data, C(h, k), for which C(h, k) failed to exceed some threshold. At the time, no method for mechanization of this clipping suggested itself. It has since been noticed that the scheme sketched in Figure 2 permits the inclusion of such clipping. Furthermore, both of the methods for choosing the clip level are easily implemented and the two implementations will be briefly described.

The first method consisted of estimating the noise present in the picture to be proportional to C(0, 0) and clipping at some pre-chosen fraction of this measurement. Since the CRT scan in Figure 2 starts near the center and spirals outward, repeating this pattern periodically, the value C(0, 0) may be estimated by integrating the amplifier output over the central portion of this sweep in order to produce the noise estimate. The additions to the block diagram of Figure 2 are obvious.

The second method was to choose a clip level such that the projected area of the C(h, k) terrain above the clip level is some pre-chosen fraction of the total aperture over which the depicted in Figure 1. In the mechanization of Figure 1 the transformed output consists of the magnitudes of the first N components of the spectrum (amplitude spectrum) and identification is accomplished by comparing observed spectra with standard spectra and choosing the best fit.

To understand the ingenious autocorrelator of [2], trace the source of illumination of a spot in the plane of the rotating disc. Since the spot is in the focal plane of the lens, its illumination comes from a certain bundle of parallel rays that enter the lens from the right, the bundle's direction corresponding to the displacement of the spot from the optical axis. Tracing this bundle farther back, we see that it has passed twice through the transparency, thanks to the plane mirror, in a way such that each ray of the bundle has traversed a pair of points in the transparency. Furthermore, since the rays of the bundle are parallel, all these pairs of transparency points correspond to the same image displacement (h, k). A smooth angular distribution of ray bundles from the diffuse light source is injected by the first half-silvered mirror (half the incident light is reflected, half transmitted).

The second half-silvered mirror similarly extracts the portion of $C(r, \theta)$ selected by the rotating sector mirror. This portion is integrated by the condenser and photoreceptor to produce $f(\theta)$.

The second process included a way to retain the information associated with the radial variation of $C(r, \theta)$. It was proposed but not pursued. However, since it affords a chance to introduce some suggestive ideas it will be described briefly. One idea behind it is the way a passive filter lies in wait for a passing waveform—the behavior of the filter may be considered a scanning of the one-dimensional function comprising the history of the signal channel. Thus if a field, in particular the autocorrelation image plane, can be scanned in such a way that the class of distortions under which we wish to regard the configuration as remaining the "same," corresponds only to variations in the time delay of the output waveform but not in its shape, then the use of a single, matched, passive filter for each pattern accomplishes painlessly the multiple cross correlation necessary for detection or recognition. (The scan of Figure 1 obviously satisfies this requirement; the waveform out of the photo receptor is only delayed more or less as the original image is translated or rotated. As the original image is magnified or shrunk it is only scaled in magnitude.) We now derive some specifications for such a scan in terms of polar coordinates (r, θ) in the autocorrelation plane. Independence of rotation evidently requires that $d\theta/dt$ be constant and that a shift in starting position of the scan not affect the output wave shape. To see what is required for independence of size, consider the effect of, say, enlarging the picture. In order to get from a portion of the auto-correlation image at one radius to a portion at another radius in the same time interval after enlargement as before, it is evidently necessary that the radial scanning speed be proportional

autocorrelation is produced. This is evidently mechanizable by a feedback circuit in which the signal that succeeds in emerging over the bottom clipper is passed through an ideal limiter and then integrated to produce a measure of the area over which the $C(h, k)$ terrain surpasses the bottom-clipping level.

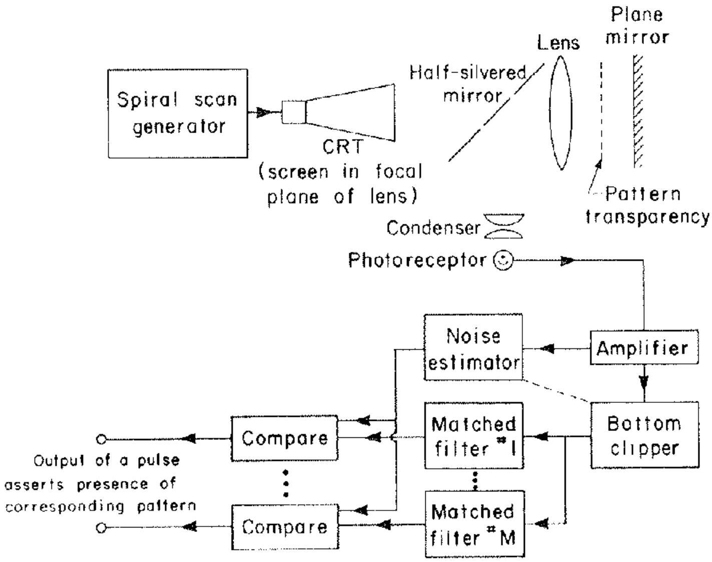


FIG. 2. A similarity-invariant pattern detection device

to radius, namely $dr/dt = ar$. A little thought shows the resulting spiral scan, $x = \exp(at) \cdot \cos bt$, $y = \exp(at) \cdot \sin bt$, cannot be independent of the starting phase in general. However, since there is a finite resolution limit, by keeping the expansion parameter, a, sufficiently small the resulting waveform can be made sufficiently insensitive to variations of the starting phase for practical purposes. (Actually $r = \exp(at)$ implies that the scan has always been going on; however, a practical scan would begin at some positive radius and repeat after reaching the perimeter of the field.) The scanning spot size should also increase in proportion to r in order to retain constant relative resolution, a logical concomitant of size-independence. Such a scan can be produced with reasonable ease on an ordinary CRT. Furthermore, it should be possible by defocussing proportionately to r to produce constant relative resolution without distorting the signal waveform. This leads to the mechanization of Figure 2, which is derived from the second method of autocorrelation function display described in [2]. Figure 2 also shows use of matched filters (rather than the spectral comparison of Figure 1) for signal recognition.

The third suggestion for dealing with the radial variation of $C(h, k)$ in a size-invariant way is to form the appropriate autocorrelation function corresponding to scale changes of the original (and hence the autocorrelation) image. The appropriate function $T(m, s)$, taking into account autocorrelation in both variables (r, θ), is

$$T(m, s) = \int_0^{\pi} \int_0^{\infty} C(r, \theta) C(mr, \theta + s) r \, d \, dr\theta.$$

Since T is a function of C, it is independent of translations of $I(x, y)$. Since C is

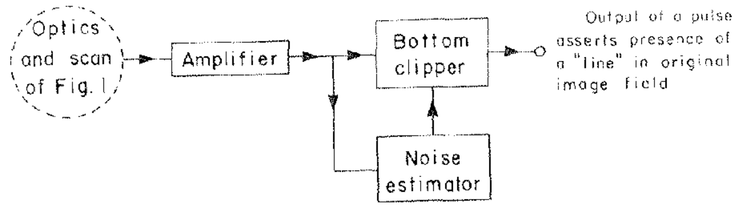


FIG. 3. A line detecting device

periodic (of period π, since $C(r, \theta + \pi) = C(-h, -k) = C(h, k) = C(r, \theta)$ in θ, T is unaffected by rotating C about $h = k = 0$ and hence by rotating $I(x, y)$ about any center. Finally, since a scale change in I results in a dilatation of C from the origin, it follows that $T(m, s)$ is only multiplied by a constant; this can be eliminated by a simple normalization to derive a final overall operation which transforms all $I(x, y)$s that are equivalent under translation, rotation and shrinkage into a single output pattern.

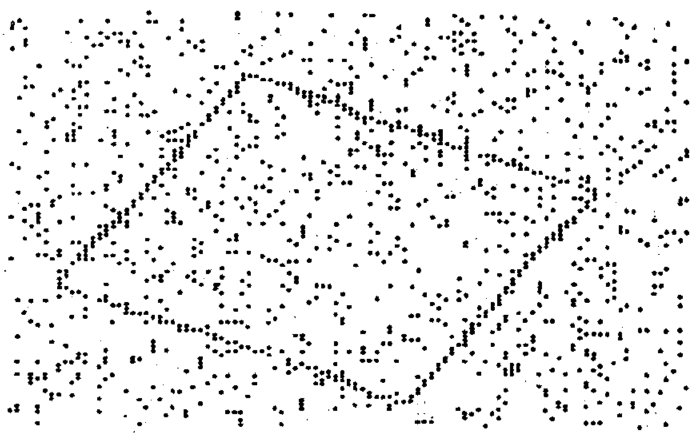
III. *Line Segment Detector*

As an example of the application of some of these notions, consider Figure 3, where a device to test transparencies for the presence of long, narrow rectangles is suggested. As can be seen, the only difference from Figure 1 is in the signal detection blocks following the photoreceptor. The scanning mirror sector acts as a low-pass filter whose impulse response matches that of an identical sector in the autocorrelation plane. Such a filter will also be reasonably well matched to the members of the set of autocorrelations of long, thin objects, the angle θ being chosen proportional to the aspect ratio expected. In Figure 3 such an object is guessed to be present if the instantaneous value of the photo pickup output is a sufficiently large multiple of the rms or the mean absolute fluctuation (whichever is used).

APPENDIX. *Computer Trial of Figure 1 Scheme*

A trial of the pattern classification scheme of Figure 1 was made on an IBM7090.[4] Because of the fairly long running time associated with digital computation of correlations and spectra, only a small number of trials took place. However, the results are felt to be significant and will be described. The number of patterns being classified was taken to be $M = 3$, the patterns being "line segment," "triangle" and "square." The transparency was assumed capable of only two densities, complete transmission ($I(x, y) = 1$) or complete absorption ($I(x, y) = 0$), in a square input image field divided into 101 by 101 resolvable spots.

[4] This work was done at Aeronutronic, Newport Beach, California, and was reported there in an "Intra-Company Communication."



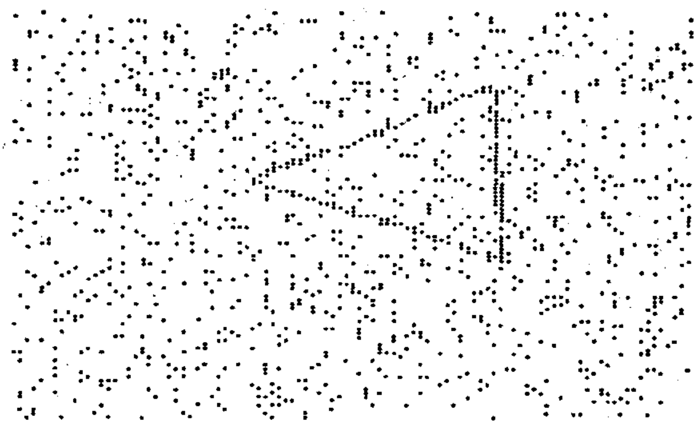


FIG. 4

Samples for attempted classification were internally generated by the following steps: (1) One of the three ideals (available as part of the computer program) is selected and then shrunk by some factor.[5] (2) The resultant is then rotated through some angle.[5] (3) The set of spots thus obtained is rounded off to integer coordinates and any duplicate spots are eliminated. (4) This resultant is shifted through two fractions[5] of the available distance to the edge of the frame (patterns were compelled to remain wholly within view). (5) Noise is sprinkled over the entire field by a process depending on a pre-set probability, p. The process consisted of two steps. First, pattern (signal) spots are removed with probability p, making use of the pseudo-random number routine. Second, at every one of the

[5] All these parameters were chosen by a pseudo-random number routine, but could have followed any other variation as easily, of course.

101 x 101 resolvable positions in the frame, a spot is similarly added with probability p. Some samples with $p = .1$ appear in Figure 4.

These samples were then processed in a way conforming to the block diagram of Figure 1. The autocorrelation, $C(h, k)$, was found and the central portion of it (a 101 x 101 square) used as the autocorrelation image field to be subjected to the sector scan and Fourier analysis. (Our optical fields are obviously round, but this detail is not important here.) Only the a-c components of the analysis were used in the spectrum matching process. The "weighting networks" were simulated by running the ideals themselves through the above described process and using the resultant component amplitudes as the weights for the corresponding patterns. (Thus the scheme in question adjusts itself to the set of patterns to be classified when suitable representatives are presented during a "learning" period.)

Using the scheme of Figure 1 in this way, three samples, at $p = .02$, were successfully identified, one of each pattern. At $p = .05$, their identification was frustrated by the large noise component in $C(h, k)$. This component is expected to be roughly uniform over the field, except as modified by the aperture autocorrelation, and for the noise mechanism used in the tests has expected value $S(h, k)p^2$, where $S(h, k)$ is the number of resolvable spots common to the original and (h, k)-shifted fields. For example, for $p = .05$ and small shifts, this is about 25. In order to eliminate some of this noise without (hopefully) too greatly disturbing the signal component of $C(h, k)$ it seemed worthwhile to use for further processing only those $C(h, k)$ which exceed some threshold. When such clipping was done near the expected level of $C(h, k)$ for noise alone, the three samples were successfully identified at $p = .05$. The noise was estimated from $C(0, 0)$, incidentally, and not taken as given through the value of p put into the program.

When $p = .1$ was tried, a different clipping rule was used: the threshold was lowered until it was exceeded by $C(h, k)$ at some pre-assigned number of points (h, k). Under these conditions and with shrinkage limited to $\frac{1}{2}$, three samples were identified correctly at $p = .1$. Probably they would also have been correctly identified under the old clipping rule at a slightly higher fixed level; however, the experiment wasn't tried.

Lest $p = .05$ or .1 seem a piddling amount of noise, it may pay to express things in terms of a signal-to-noise power ratio. A reasonable measure for input S/N is simply the ratio of the number of black spots representing the pattern to the number representing the noise. For the trials at $p = .05$, for example, $E\{N\} = (101)(101)(.05) \approx 500$. The patterns actually used contained from $S = 70$ to $S = 190$ spots. Thus S/N varied from -8 to -4 db in the successful trials at $p = .05$. For $p = .1$ the range was, of course, -14 to -10 db. For the trials at $p = .02$, without any autocorrelation clipping, the corresponding S/N ratios are -4 to 0 db. No other comparable results seem to have been revealed elsewhere for the problem in question—pattern recognition independent of location, orientation and size and in the presence of equal or stronger disturbing noises.

REFERENCES

1. Elias, Peter; Grey, David S.; and Robinson, David Z. Fourier treatment of optical processes. *J. Opt. Soc. Am. 42* (Feb. 1952), 127-134.
2. Kovasznay, Leslie S. G., and Ali Arman. Optical autocorrelation measurement of two-dimensional random patterns. *Rev. Sci. Instr. 28* (Oct. 1957), 793-797.
3. Clowes, M. B., and Parks, J. R., "A new technique in automatic character recognition. *Comput. J. 4*, (July 1961).