INTRODUCTION

The design of a system for recognition of patterns or objects with provision for size and rotation invariance\(^1\) will be described. This system is also self-learning in that a program of remembering a pattern by "seeing" that pattern is incorporated in the system. The process of "free learning" which is defined here as abstracting from the results of numerous inputs and outputs to arrive at a more general recognition class is also part of our design. We shall first briefly describe the properties incorporated in our design, and then outline the system.

The nature of recognition is such that the image of each object belongs almost to an infinite class in the sense that a retinal or photoreceptor matrix array can be stimulated in a very large patterns or objects with provision for size and rotation invariance. Therefore the first design task is the recognition of the same object with size and rotation invariance. The second important design principle is that there be a recognition of a set of objects which are the same in the sense that a single output is required for all members of the set even though the members all vary somewhat image-wise (e.g., all A's are members of one set even though not drawn in the same way). Third, the coded differences between different sets of images (A's, B's, C's, etc.) must be maximized in order to also maximize the permissible variation within a set. These three guidelines for design are to be incorporated in the system logic so as to permit the recognition procedure to interpolate internally by a type of self-learning process.

No learning can take place without a feedback or correcting signal. When humans learn to read, they are given a simultaneous set of inputs, i.e., they are shown letters and asked for a specific response to those letters. After a time, the human learns to internally program his learning. That is, by trying out various responses, he stimuli the desired response is reached using a nebulous error correcting signal. We can design in a much simplified manner a homothetic transformation having a unique ordinary point at its center. For rectangular coordinates the transformation is simple:

\[
\begin{align*}
    x' &= ax + c_1 \\
    y' &= ay + c_2
\end{align*}
\]  

(1)

where \(a\) is the is the dilation scale factor and is not zero or one. The transformation has a center given by:

\[
\begin{align*}
    x &= c_1 (1 - a)^{-1} \\
    y &= c_2 (1 - a)^{-1}
\end{align*}
\]  

(2)

The set of dilation transformations does not form a group since it is not closed under multiplication, nor is there an identity transformation.

If we consider an image of a figure on a plane with border coordinates \(x_1, y_1\) with the centroid of the image initially centered, the constants \(c_1\) and \(c_2\) may be set to zero in the transformation. The unique point of the dilation is then \((0,0)\) which is the center of the image.

The dilation of an image occurs externally to the photoreceptor system, and may be considered
as a projection of an image as the projector is brought closer to the receptors. Similarly, of course, an image size reduction may be interpreted as a projector withdrawal from the receptors.

The physical problem consists of arranging the image reception system so as to provide the same digitized image code for the image regardless of dilation (except that the image must be within the receptive area of the photoreceptors). A simpler dilation transformation than \( t \) may be utilized if we work in polar coordinates. The corresponding polar dilation is

\[
\begin{align*}
\rho &= a \rho' \\
\phi &= \phi'
\end{align*}
\]

Therefore the simplest arrangement of photocells is a polar array. The constant of dilation \( a \) is not constant for all projected images, but takes all possible values up to the maximum value

\[
m_{\text{max}} = \frac{r_{\text{max}}}{r_{\text{p}}}
\]

where \( r_{\text{p}} \) is the minimum radius of a resolved image centered on a polar arrangement of photocells, and \( r_{\text{max}} \) is the maximum possible radius for a resolved image contained in the photocell array.

### PHYSICAL DESIGN

A physical embodiment of the general principles is shown in Figure 1. The photocells are arranged in a polar array to create an electronic "retina" which provides for a set of pulses which can be logically connected so as to provide a size-invariant image code. Each segment in a sector of the polar arrangement of photoreceptors contains four or more independent photocells which are connected to provide an output only if an image border is projected upon a segment. The method of combining photocells four at a time is shown in Figure 2. The logical (Boolean) analysis of requiring one, two, or three photodectors to be in darkness (or illuminated) but not all four to be equally stimulated to obtain an output, is

\[
(a+b+c+d) \cdot (a' \cdot b' \cdot c' \cdot d')
\]

where \( a, b, c, d \) represent the binary state of each cell, a bar represents "not", the dot is for logical "and", and the + is for logical "or".

Using this model of a bridge circuit, each photoreceptive border receiver area must become larger with increased distance from the center of the polar array of Figure 1. There are several methods of accomplishing the increased reception area. If solid state photoreceptors are utilized, the required area is easily arranged. However, another approach utilizes glass (also quartz, etc.) fibers to guide the light paths to the photocells. This permits a very small area to be resolved. It is particularly useful to have very small photoreceptor areas in towards the center of the polar array to decrease the area of the central blind spot, and fiber light guides may be invaluable for application here. There is, however, a more important reason for utilizing fiber optical wave guides to collect light and direct it to the photoreceptors. The reason is economy of electronic elements with essentially no loss of resolution. Away from the center of the polar array of photodetectors, the segments are large. Only one pulse output signal must come from each segment. In order to "see" fine points in the outer segments, individual photosensitive elements must be smaller than the point. By using thin glass fibers leading more or less randomly (within a segment) to four (or eight) bridge-connected electronic photoreceptors, the resolution of a small signal (narrow line) is good, while the number of electronic components is relatively small. The effect of this arrangement is to permit even one stimulation of a single light fiber to cause a signal to emerge from one polar segment. The more or less random connections of the optic fibers to the photocell bridge lead to a possibility of cancelling a set of light spots which stimulate the four photodetectors of the bridge equally. Due to the lack of well-organized connections, the probability of such a pattern occurring is almost nil.

It may be interesting to note that the human retina may be organized in an analogous way. Near the center of the retina only a few photosensitive retinal elements (rods) are interconnected neurally before being connected to a fiber of the optic nerve. The rods (or cones) towards the periphery of the retina are interconnected sometimes hundreds at a time before being connected with one optic nerve fiber\(^5\).

We have discussed the arrangement of the photocells into bridges of fours with each set of four photoreceptors filling a segment of the polar plot of Figure 1. The next topic to be considered is the logical arrangement of output signals from the polar array. The output is most readily and conveniently manipulated by a general purpose digital computer. Even though many of the computer operations will not be utilized, the availability of such computers is an economical gain.

### THE PHOTOMATRIX

The photodetectors are energized by a pulsed voltage source for a short time -- the order of a microsecond -- just long enough to exceed the response time of the light detectors. The repetition rate for this activating pulse is relatively slow (about 30,000 pulses per second). The repetition rate of \( 1/r_{\text{p}} \) of the response time, where \( r_{\text{p}} \) is the total number of circular divisions of photoreceptor areas, is utilized to code the image for recognition or memorization prior to the appearance of a second image.

The outputs from the polar segments are connected as shown in Figure 3. Each output goes through a unilateral conduction element (a diode) and a delay (of about a microsecond which is about the receptor response time) before being attached to radially neighboring photoreceptors. One has a choice of directing the output pulse path either radially outward or inward, and we have selected, for the present system, to arbitrarily use radially outward propagating signal pulses.
Let us now observe the effect of our logical and geometrical circuit combination. Suppose an image appears on the photosensitive array shown in Figure 1. The borders of the image which fall upon segments of the array activate the photoreceptor bridges (Figure 2) lying in those segments (when a positive pulse occurs) and a set of pulses representing the borders travels outwards along the radial connecting lines shown in Figure 3. The activating pulse also starts a "clock" which determines the position of all pulses on a time scale. We have, then, thirty-six radial lines either carrying pulses or not with the pulse position (in time) indicating the borders of an image. We may consider the digitized image representation to be a matrix of binary elements with the rows corresponding to the radial output lines and the columns corresponding to the time scale (each column is separated by one delay period or clock time unit).

**THE MATRIX REPRESENTATION OF AN IMAGE**

The image is converted to a set of radially outward traveling pulses by the reception system described above. We now wish to describe the combination of pulses to represent the image for both learning and recognition purposes.

It is convenient to number the radial pulse propagating lines in a clockwise manner. The most vertically upward sector then of Figure 1 is called radial number 36 and the sectors to the right are radials 1, 2, 3, etc. If the radials are given rows in a matrix of binary numbers then the time sequence of pulses in each radial determines the position of the units in the matrix. The matrix size is given by: the number of radial lines = number of rows of the matrix, and the number of delay units = the number of columns of the matrix. We assume here that each delay unit is equal between segments of the photosensitive array and that a unit delay follows the outermost segment.

Having described the image coding scheme, we will now illustrate how images which are centered upon the polar photosensitive array provide a binary matrix code. Consider images of alpha-numeric (letters and numbers) form first. Suppose that the width of the lines of these letters are larger than the diameter of a photoreceptor element (e.g., a glass fiber conductor), and also that the image is of a size suitable for total inclusion in the photosensitive array.

For heuristic simplicity consider on "0" first. The figure is imaged, the photocells energized by a pulse, and the pulses representing the "0" travel down the radial delay lines in synchronism with the digital clock. In conformity with common practice, absence of a pulse denotes a zero, and a pulse denotes unity. The matrix denoting the "0" code is then obtained by taking the number one radial "word" (set of pulses) as the first row of the matrix followed by the number two radial word, etc. The form of the matrix is indicated in abbreviated form as

\[
\begin{bmatrix}
0 & 0 & \ldots & 0 & 1 & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 1 & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 1 & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 1 & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 1 & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\end{bmatrix}
\]

It may be noted that the zeros both to the left and right of the columns of ones provide information about the size of the image, but nothing regarding shape. Therefore we use a floating decimal point or similar type operation to eliminate all columns of zeros to the left and the right of the first columns from the left and from the right which contain units. We may retain information regarding the number of columns dropped as an indication of image size if we chose, but for the moment, that is not important. The reduced matrix is then

\[
\begin{bmatrix}
1 & 0 & \ldots & 0 & 1 \\
1 & 0 & \ldots & 0 & 1 \\
1 & 0 & \ldots & 0 & 1 \\
1 & 0 & \ldots & 0 & 1 \\
1 & 0 & \ldots & 0 & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\end{bmatrix}
\]

Now, the columns containing entirely zeros do not provide much information. These only indicate the "thickness" of the image. Nonetheless, the thickness is an invariant of an image; due to the photoreceptor geometry, a dilution of the image provides the same thickness in the matrix representation. Therefore, we choose not to operate upon the "internal" columns containing all zeros. The reduced matrix M2 is called the "canonical matrix" for "0".

We next consider the redundancy of information contained in matrix M2. It is apparent from a cursory examination of the matrix that the outer shape (outer border) of a figure is represented primarily by the change of position of the occurrence of units (i.e., nonzero elements) in the right hand columns. Therefore, it is reasonable to utilize the difference between row positions of the right hand units in the canonical matrix. For the M2 matrix the difference is given by a string of zeros since all of the units occupy the same row. The canonical outer border is then:
This word is stored in the memory as a representation of an 0 or zero when the system is in the learning mode.

It will be useful to next give an example of a less symmetrical pattern such as a line drawing of the letter D. This letter is shown "centered" in the photomatrix of Figure 4. The canonical matrix of the letter is shown in (M3).

<table>
<thead>
<tr>
<th>Sector No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ... 0 0 0 1 1 0 0 0 0 0 0 (1) ... ]</td>
</tr>
<tr>
<td>[ ... 0 0 1 1 0 0 0 0 0 0 0 ... ]</td>
</tr>
<tr>
<td>[ ... 0 0 1 0 0 0 0 0 0 0 0 ... ]</td>
</tr>
<tr>
<td>[ ... 0 1 1 0 0 0 0 0 0 0 0 ... ]</td>
</tr>
<tr>
<td>[ ... 0 1 0 0 0 0 0 0 0 0 0 ... ]</td>
</tr>
<tr>
<td>[ ... 0 1 0 0 0 0 0 0 0 0 0 ... ]</td>
</tr>
<tr>
<td>[ ... 1 0 0 0 0 0 0 0 0 0 0 ... ]</td>
</tr>
<tr>
<td>[ ... 0 1 0 0 0 0 0 0 0 0 0 ... ]</td>
</tr>
<tr>
<td>[ ... 0 0 0 1 1 0 0 0 0 0 0 ... ]</td>
</tr>
<tr>
<td>[ ... 0 0 0 0 1 0 0 0 0 0 0 ... ]</td>
</tr>
<tr>
<td>[ ... 0 0 0 1 1 0 0 0 0 0 0 ... ]</td>
</tr>
<tr>
<td>[ ... 0 0 1 1 0 0 0 0 0 0 0 ... ]</td>
</tr>
<tr>
<td>[ ... 0 1 0 0 0 0 0 0 0 0 0 ... ]</td>
</tr>
<tr>
<td>[ ... 0 1 0 0 0 0 0 0 0 0 0 ... ]</td>
</tr>
<tr>
<td>[ ... 0 1 1 0 0 0 0 0 0 0 0 ... ]</td>
</tr>
<tr>
<td>[ ... 0 1 0 0 0 0 0 0 0 0 0 ... ]</td>
</tr>
</tbody>
</table>

The canonical word representation of the letter D found by taking the difference between row positions of outer units is:

\[-1, -1, 0, -1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, +1, 0, -1, 1, 1, 0, -1, 0, -1, 0, -1, 0, -1, 0, +1, 0, +1, +2, 0, +1, 0, -1\]

This is the canonical word which is stored as a memory representation of the letter D shown in Fig. 4. Of course, the digital manipulation to obtain this word is simply programmed into the machine logic and one is never aware of this representation. Once a "D" has been exposed to the photomatrix of the machine while in the learning mode, the canonical word is automatically calculated and stored in the memory.

**READING THE INPUT IMAGES**

Suppose that after a number of such words are memorized, the machine is in the reading (or recognition) mode. Images of letters (or words) presented to the photomatrix array are converted to canonical words which are then compared to the memorized words. The comparison may be performed in a number of ways, for example by subtraction, by sieving, or by some other logical operation. For simplicity of the heuristic design, we will assume here that the comparison is by subtraction between the memorized words and the word to be recognized. The system looks for the absolute value of the difference between all of the stored words in a certain subclass of the entire collection of the memory. (This subclass is determined by examination of canonical words for similar symmetry properties.) The initial search is for a difference of perhaps ten units or so. This would be a poor match. Suppose that three such matches at three memory addresses are found. The question now arises as to whether these three matches are members of the same class of symbols. That is, are they all memorized D's with various distortions? This question is answered by a logical procedure of comparing output signals from the three addresses. Suppose they are not identical. Next a comparison is run among the four words again, allowing for perhaps an absolute difference of five units in this comparison. Suppose that there are now two matches. A third run allowing for an absolute difference of three units or so will usually narrow down the match to one. The machine may now print out the address of the memory word which was the best match. Usually the address has been coded in the learning process to correspond to the letter D.

Now the system has been programmed to re-examine the three memory words which matched within units of the input word. Suppose that of these three words, two correspond to the same output (let's say "D"). The system retains information about this redundancy. Such redundancies are sometimes useful, but too many will be uneconomical of memory storage space. Therefore redundancies are programmed for removal either when (1) the memory is overloaded, or (2) a canonical word has considerable overlap with a
canoncal word belonging to another set. The examination of the memory for nameful redundancies takes place during the reading (matching) process.

If no suitable matching (recognition) can be accomplished for a symbol, the machine changes its state from a reading mode to a learning mode and stores the canonical form of the input image. Unless an output operator now gives this symbol a specific "name" so that the system may print out the desired name of this image, the machine will, when reading this symbol, simply print out a number corresponding to the address of the stored canonical word.

ROTATION INVARIANCE

In this system, an image can also be examined for matching with memorized canonical words when there is relative rotation of the images. The effect of a rotation is to change the row order of the canonical matrix. That is, a rotation of an image by ten degrees clockwise is identical to placing the top row of the canonical matrix on the bottom of the matrix. In a similar way, any amount of image rotation is equivalent to a corresponding amount of canonical matrix row rotation. Thus in the recognition process, it is quite simple to look for a "match" with a certain amount of rotation added to the canonical word form.

SUMMARY

The system is composed of an electronic shutter, a photoreceptor matrix (the retina), a centering system, a pulse timer, and a digital computer with a specific program. When the machine is in the learning mode, images appearing on the retina will be centered and transformed (with size invariance) to a canonical matrix which in turn may be reduced to a canonical word stored in the computer memory.

The images need not be simple alphabetical patterns; they may be human faces, navigational guileposts, or any other sort of image. For example, one may utilize an extension of our ideas to identify human faces or navigate by terrain recognition.

After a sizable memory is obtained, the machine may be put into the read mode. It is to be noted that the machine is usually in the read mode. Each incoming image is tested for a certain degree of matching with previously stored patterns. If the approximate matching result indicates that the image is not "identified" the machine will automatically go into the learning mode and store the canonical word representing the image.

The canonical word (or words for complex images) is size invariant in the sense that a large or small image will result in the same representational code. Rotation invariance occurs through testing the representation for rotation. Each canonical matrix representation can be tested for left hand rotation by removing matrix rows from the top and putting these on the bottom. Right hand rotation reverses this matrix manipulation.

The matching procedure between the memory and the image representation allows for a certain degree of "smoothing". First of all, "similar" patterns which are distorted relative to each other will occupy different memory positions. Thus there is ability to recognize misshapen patterns through utilizing additional memory space. Secondly, the matching proceeds through an elimination of mis-matches and narrowing of possibilities. The difference between the image representation and the stored canonical forms is permitted to be large (a large number of differences) at first. Then, as the possibilities are lessened, a better match is attempted. The mathematical procedure of smoothing is programmed, that procedure which can be programmed on any electronic digital machine (although some machines are easier to employ than others). It should be noted that the conventional decision procedure of electronic computers (i.e., "is A > B?" is more powerful than it first appears. For example, by adding a predetermined number x to A or B one can manipulate the logic to permit a decision as to whether A is x units larger or smaller than B. This then, is a smooth match between A and B.

The machine is also programmed to correct for overlapping sets of figures. For example, the letters U and V have rather similar canonical representations. Consequently the maximum amount of smoothing permitted for these letters will be automatically found by machine logic with allowance for the fact that very little overlap in the recognition should be tolerated. The significance of this in the recognition procedure is that the smoothing of U and V in the reading procedure will be found to be less than for other patterns which are not as similar. Such problems are entirely solved by the machine program. The determination of smoothing values are found by looking for the maximum which separates out memory units having different output signals. In other words, the machine will continue the matching procedure as long as any of the matching memory words have different outputs, but will stop when the matching memory addresses (and they may well be multiple) correspond to the same set (same output).

Since each pattern can have considerable malformations, there may be a number of memory locations corresponding to a particular output. During the course of reading, the machine is to be programmed to recognize these redundancies and eliminate them whenever they are not required to prevent overlap with other patterns. It may be well to explain this more fully. If a large smoothing is possible for some particular pattern without confusing different sets of patterns (patterns with different outputs) then there is no need for memorizing many forms of that particular (since the forms with malformations are recognized through smoothing). Consequently, the machine program eliminates stored patterns which
do not help to isolate a recognition set. This program is needed if we are to utilize a machine with a small memory capacity.

Finally, there are a number of ramifications planned. Recognition of words, people, assembly line parts, etc., requires more memory capacity than letters, but the procedure is the same. It is perhaps simplest to store the canonical matrix rather than canonical words as more information is then handled in a uniform manner. With the additional information, redundancy increases. Consequently, some of the recognition problems are lessened. For example, distinguishing between U and V is much simpler when these are contained in words than when they stand alone.

ACKNOWLEDGEMENT

It is a pleasure to thank Peter M. Kelly, Lloyd Lambert, and C. L. Vanlass of the Aeronutronic Physical Electronics and Bionics Division for their interest and helpful criticism.

REFERENCES


FIGURE 2

PULSE ACtIVATION SOURCE

PULSE OUTPUT FOR BORDER STIMULI