"AUBERT PROCESSING" AND INTELLIGENT VISION

Stewart W. Wilson
Research Laboratories
Polaroid Corporation
Cambridge, Mass. 02139

In 1857, Aubert and Foerster(1) reported that the off-axis angle at which an indirectly viewed letter is just recognizable is proportional to the letter's size. That is, doubling the size of a letter makes it recognizable at twice the eccentricity, and so forth. In symbols, if A is the letter's visual angle and E is the angular eccentricity at which it is barely recognizable, then $E = cA$ with c a constant dependent on the letter's form and the experimental circumstances.

This result suggests that in humans, in contrast to most Intelligent Vision systems(2), "processing power per unit area" is not spread uniformly over the visual field, but is instead proportional to $1/E^2$. To motivate the inverse square, assume that the same letter, whatever its size, requires the same "total processing power", P, for recognition. Then, if a letter of size A is just recognizable at eccentricity E, the "processing power per unit area" at that point must be $p = kP/A^2$ with k a constant. But since $E = cA$, it follows that p is proportional to $1/E^2$.

The array of Fig. 1 shows a structural interpretation of this relationship. Each ring about the center has the same number of circles and there is a constant factor of overlap from ring to ring(3). Since the area of a circle is proportional to the square of its distance from the array center, the number of circles per unit area is proportional to $1/E^2$. This makes it tempting to think of each circle as the receptive field in retinal space of a standard unit of image processing machinery. Then the number of such standard units per unit area--or, assuming additivity, the processing power per unit area--

would be proportional to $1/E^2$. A letter superimposed on the array at eccentricity E_1 would, if doubled in size and moved out to $2E_1$, be "examined" in both cases by the same number of processing units. Recognizability would presumably not change, consistent with Aubert and Foerster's result.

The existence of standard processing units whose receptive field diameters and spacings increase linearly with eccentricity has been shown physiologically in monkeys, notably through work of Hubel and Wiesel(4). It therefore seems important to ask whether visual structures having the form of Fig. 1 ("Aubert processing arrays") could be relevant to the development of devices for Intelligent Vision.

To investigate this question we shall examine an image processing model in which the circles of Fig. 1 are each the "data field", or retinal area over which luminance input is gathered, of one of a set of standard processing machines, or "Aubert Units" (AU's). The AU's are functionally identical and only differ in the size of their data fields. The AU performs a standard set of computations on its data field input and it outputs a standard set of signals indicating the results of those computations. If two data fields differ in size by a factor of r and their luminance inputs have the same spatial pattern except for a scale factor of r, then the output messages from the associated AU's will be identical. Without committing to any specific set of output messages, the AU's might usefully compute average input brightness (in wavebands), the presence or absence of certain elementary forms, etc.

A convenient way to handle the AU output messages systematically is to imagine a scanning process beginning with AU's at the center of Fig. 1 and going outward in a circle. This has the attractive property that for different sizes of the same object, similarly fixated, the output message ensembles will be identical except for a time-shift. Similarly, rotating an object about the fixation point would only rotate the messages with respect to the AU's. The message ensemble itself would be unchanged. In contrast, however, a given object produces radically different message ensembles under translation of the fixation point; we shall return to this apparent difficulty.

A related way to process the messages is to imagine a mapping in which the AU's associated with data fields on "rays" and "rings" of Fig. 1 are arranged in the rows and columns, respectively, of a rectangular, matrix-like space as shown in Fig. 2. Here, the ring closest to the center in Fig. 1 maps into the left-hand column of the matrix. The matrix's top row is the "12 o'clock" ray, its middle row the six o'clock ray, and on clockwise to the bottom row which is the ray just before 12 again.

Schwartz(5) has pointed out that neurophysiological data imply that the primate visual system maps position from retina to primary visual cortex in essentially the manner just described. He shows that the mapping has the form $w = \ln z$, where $w = u + iv$ is position in cortical space and $z = re^{i\varphi}$ is position in retinal space. Then $u = \ln r$ and $v = \varphi$. This is consistent with the derivation from Aubert and Foerster. In Fig. 1, the distance ΔE from one data field center to the next is proportional to E itself which implies that the column index in the above matrix is logarithmic in E. Similarly, the fact that all rings in Fig. 1 have equal numbers of data fields directly implies that the matrix's row index is linear in polar angle.

With the AU's and their output messages arranged in the "cortical" matrix of Fig. 2, it is evident that the circular scan of retinal space becomes a vertical bar scanning the cortex from left to right. Fig. 3 shows cortical "images" of a clown seen at two distances differing by a factor of three. (The pictures(6) show just gray scale values but of course the cortical activity would be a superposition of distributions of all message types.) The clown's image, though "distorted", is of constant size and shape. Fig. 3 also shows the result of rotating the clown through 45 degrees; again, cortical size and shape remain the same. Retinal scale change and rotation only alter the position of the cortical image. One might thus say that Aubert processing disentangles the form or pattern of an object from "accidental" circumstances (size, orientation) of its relationship to the viewer. Fig. 4 illustrates this for a texture. There, a uniform grid is seen at distances differing by a factor of two, and in rotation. The cortical images are the same except for a shift.

Templates have tended not to be used for sophisticated Intelligent Vision because they are easily "fooled" by size and orientation differences. The invariances just adduced suggest, however, that templates may find their true utility in "cortical" space. One can imagine the scanning bar referred to above progressing from left to right and emitting successive column message sets which are compared with a library of column message set templates. Each such template might in fact be the input side of a production rule. The output would be an internal message (to permit dependencies from column to column), an action command (perhaps to move the eye to see a place better), or a recognition or other decision, etc. The templates might arise through experience and a reward structure, in an adaptation of ideas of the sort suggested by Holland(7). In our laboratory we are constructing an "animaton" which will incorporate Aubert processing, template matching, and physical movement; of course the concepts are presently tentative and experimental.

Earlier it was noted that the Aubert-processed image changes radically with shift of fixation point. This would appear to cast doubt on the usefulness of a template system. However, off-center objects tend to turn into simple "blobs" under Aubert processing, just as they tend to do in human vision. If we relax the requirement that off-center objects be immediately recognizable, we can imagine general or "crude" shape templates serving to bring peripheral objects to the center. Once there, an object would fire the standard and detailed "center-view" productions. A similar strategy might be helpful in handling objects that are centered but have a non-standard orientation.

The essential attribute of Aubert processing is its independence of scale, achieved in effect by applying all scales at once: resolution is in principle infinite at the center and coarsens radially outward. Thus scale, resolution, or "grain" problems which can accompany linear vision techniques do not arise. At the same time, a more meaningful kind of resolution, independent of the perceiver-object relationship, becomes possible. Note in Fig. 1 that the number R of AU's per ring is a free parameter: if R is small, all centered objects, regardless of size, are viewed coarsely; if R is large, all objects are seen in detail. R, a property of (and potentially under the control of) the perceiver, determines the resolution with which objects themselves, not images of them, are analysed. This seems a powerful feature of the Aubert design.

[January 30, 1981]

NOTES AND REFERENCES

1. Aubert and Foerster's experiments are described in H. v. Helmholtz, Treatise on Physiological Optics, James P. C. Southall, ed., Dover, N.Y. 1962, Part II, pp. 39-42.

2. A significant exception is the "retina" used by Funt in his WHISPER program; it is the same in certain respects as the structure discussed here. See B. V. Funt, "Problem-Solving with Diagrammatic Representations", Artificial Intelligence, vol. 13, no. 3, May 1980, pp. 201-230.

3. The innermost ring (not shown) has a finite, though very small, radius.

4. D. H. Hubel and T. N. Wiesel, "Uniformity of Monkey Striate Cortex: A Parallel Relationship between Field Size, Scatter, and Magnification Factor", Journal of Comparative Neurology, vol. 158, no. 3, December 1, 1974, pp. 295-305.

5. E. L. Schwartz, "Spatial Mapping in the Primate Sensory Projection: Analytic Structure and Relevance to Perception", Biological Cybernetics, vol. 25, 1977, pp. 181-194.

6. In this computer-generated mapping there are 64 AU's per ring and per ray. Fixation is on a point between the clown's eyes.

7. J. H. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, 1975.

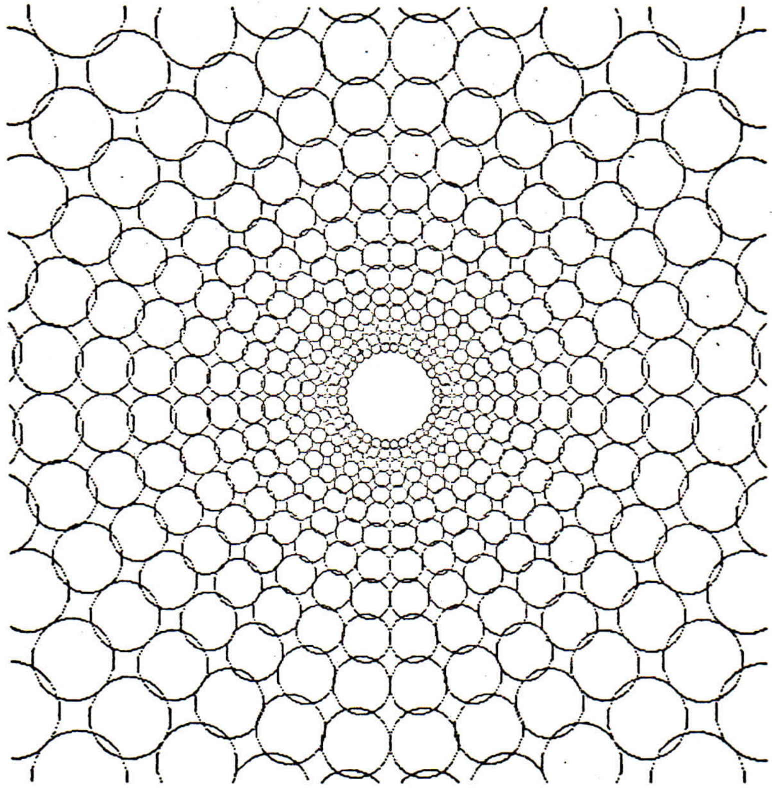

Fig. 1

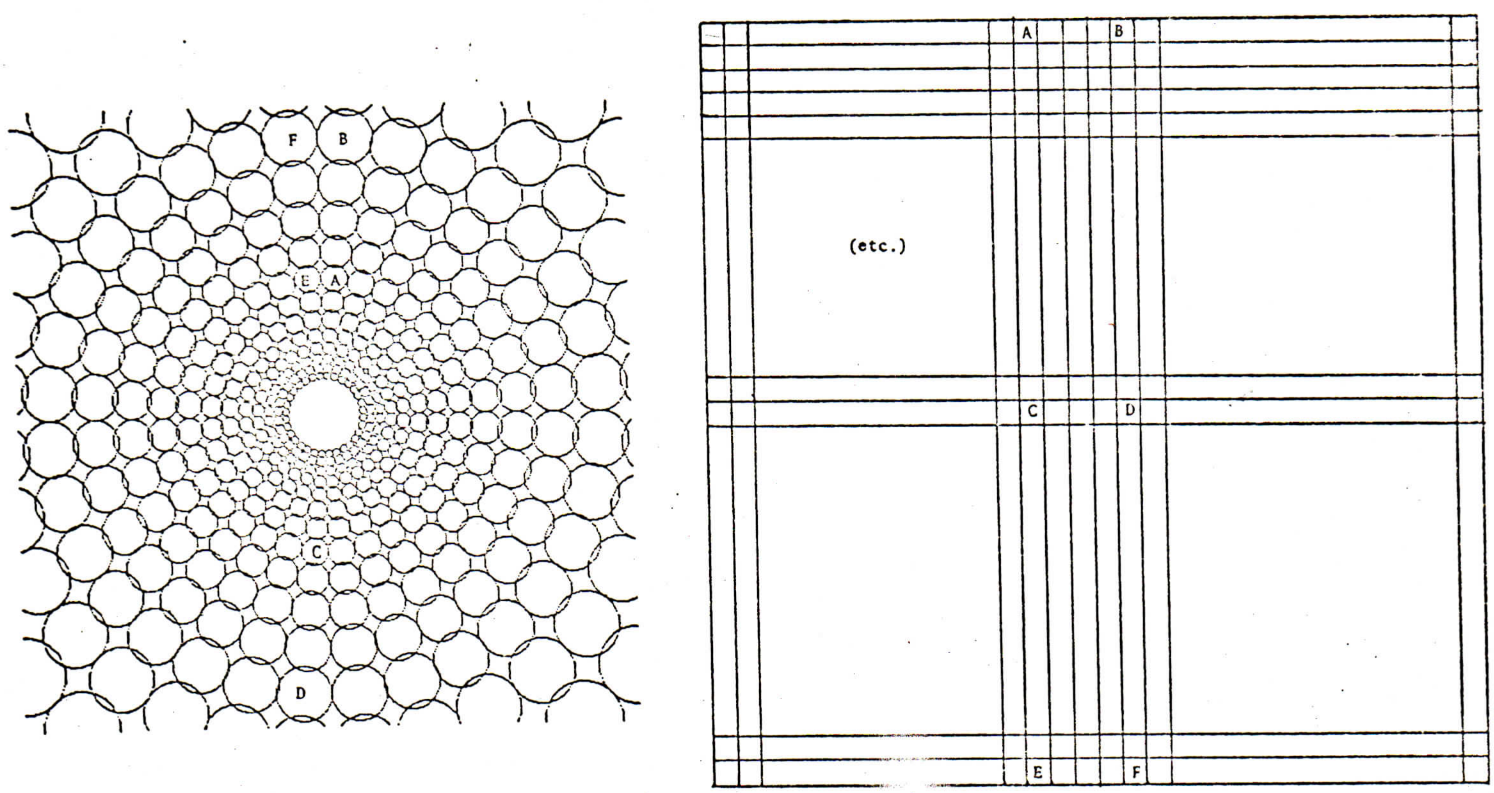

Fig. 2 Aubert Mapping

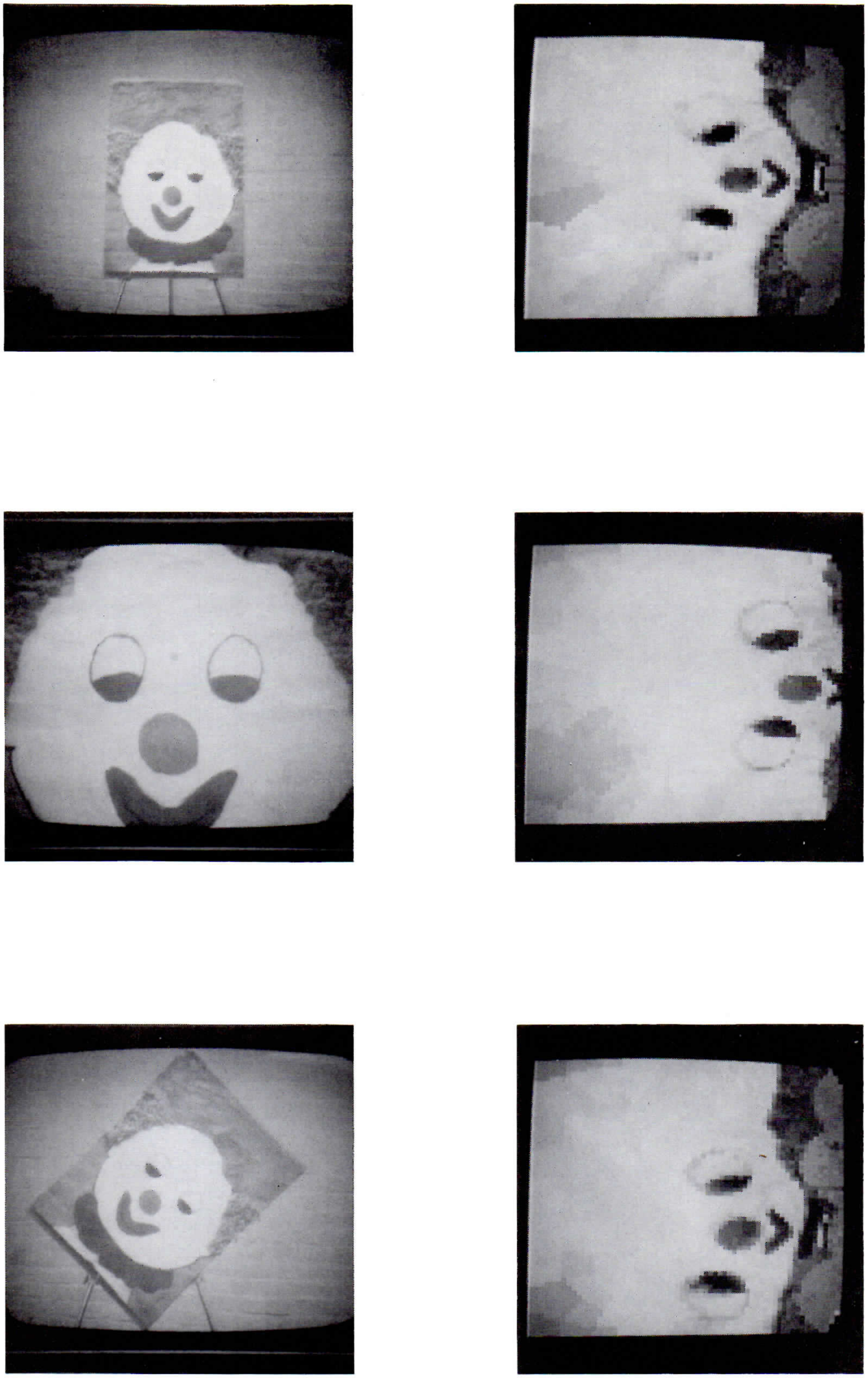

Fig. 3

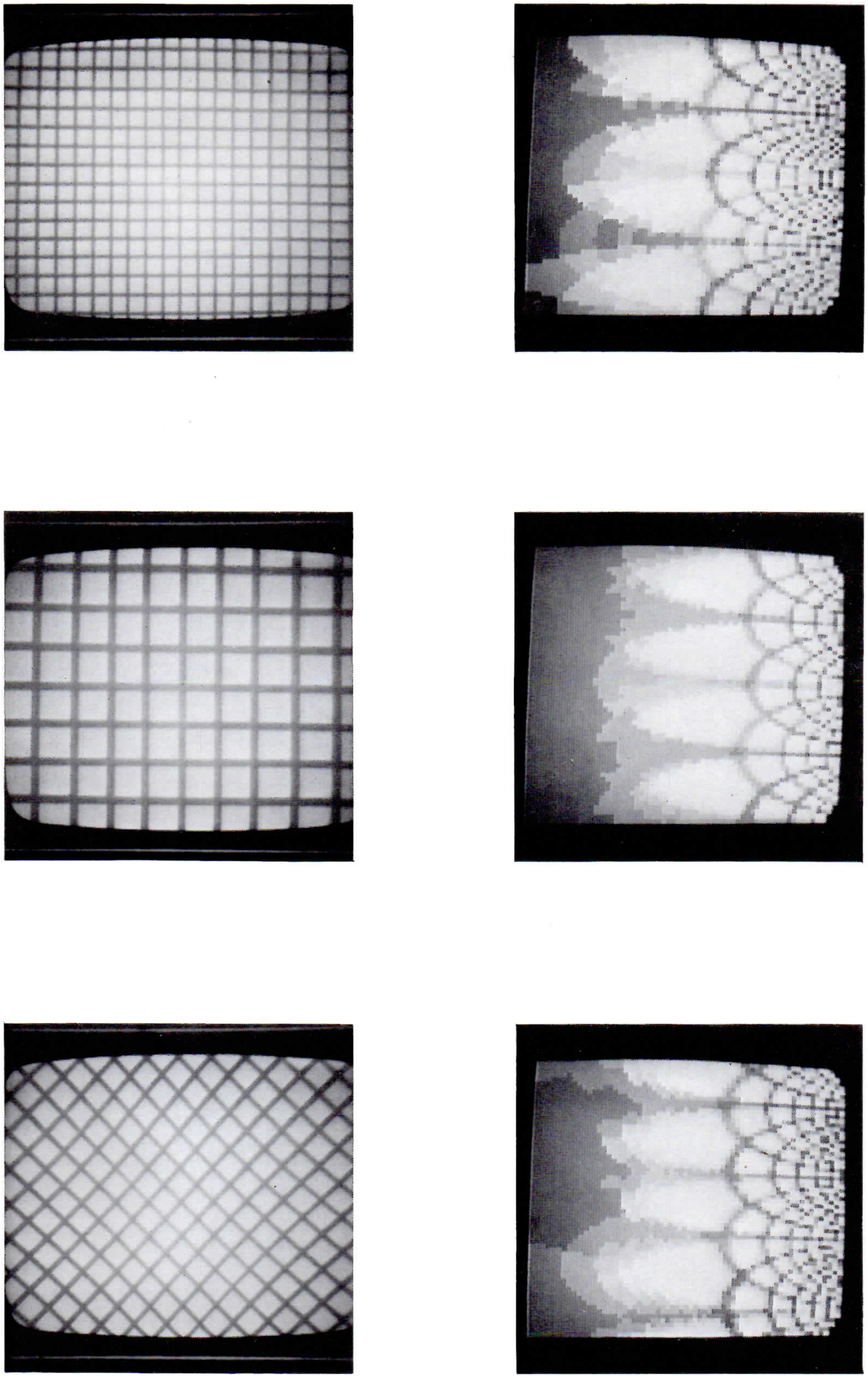

Fig. 4