ADAPTIVE "CORTICAL" PATTERN RECOGNITION

by

Stewart W. Wilson

Rowland Institute for Science, Cambridge MA 02142

ABSTRACT

It is shown that a certain model of the primate retino-cortical mapping "sees" all centered objects with the same "object-resolution", or number of distinct signals, independent of apparent size. In an artificial system, this property would permit recognition of patterns using templates in a cortex-like space. It is suggested that with an adaptive production system such as Holland's classifier system, the recognition process could be made self-organizing.

INTRODUCTION

Templates are generally felt to have limited usefulness for visual pattern recognition. Though they provide a simple and compact description of shape, templates cannot directly deal with objects that, as is common, vary in real or apparent (*i.e.*, imaged) size. However, the human visual system, in the step from retina to cortex, appears to perform an automatic size-normalizing transformation of the retinal image. This suggests that pattern recognition using templates may occur in the cortex, and that artificial systems having a similar transformation should be investigated. Properties of the retino-cortical mapping which are relevant to pattern recognition are discussed in the first half of this paper. In the second half, we outline how an adaptive production system having template-like conditions might recognize patterns that had been transformed to a "cortical" space.

THE RETINO-CORTICAL MAPPING

Recent papers in image processing and display, and in theoretical neurophysiology, have drawn attention to a nonlinear visual field representation which resembles the primate retino-cortical system. Weiman and Chaikin [1] propose a computer architecture for picture processing based on the complex logarithmic mapping, the formal properties of which they analyze extensively. They and also Schwartz [2]

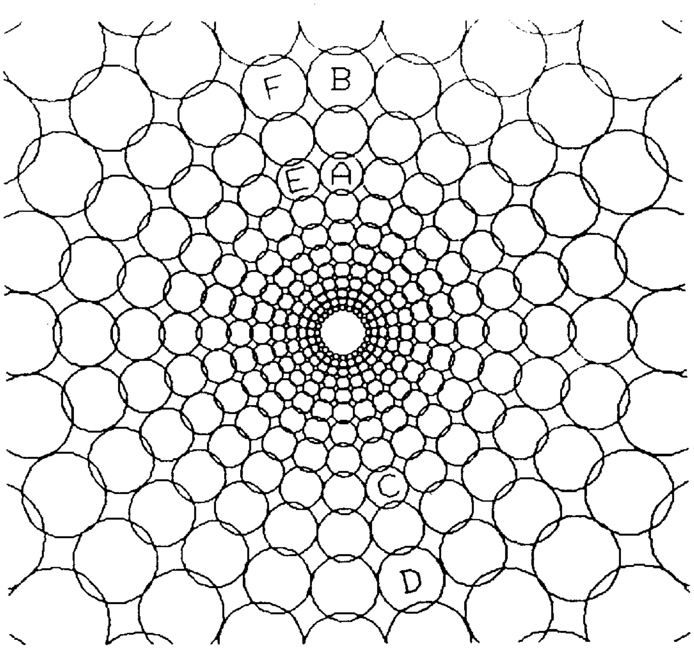

Figure 1. "Retina" consisting of "data fields" each connected to an "MSU" in the "cortex" of Fig. 2.

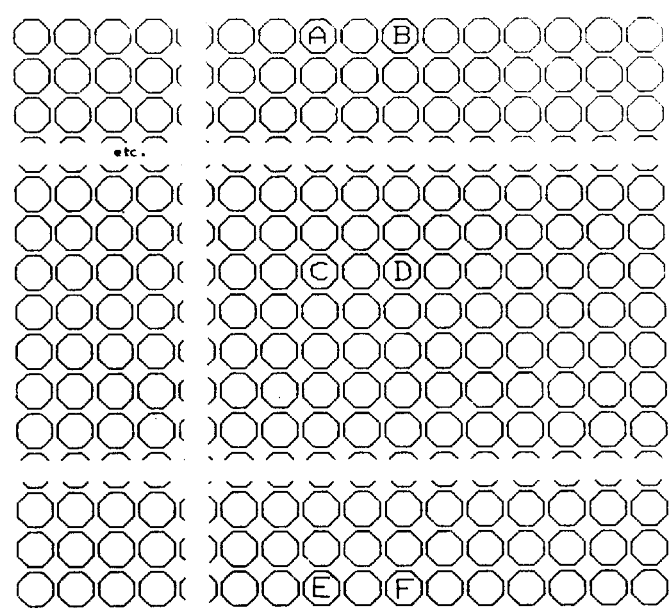

Figure 2. Each MSU receives signals from a data field in Fig. 1. Letters indicate connection pattern.

present physiological and perceptual evidence that the mapping from retina to (striate) cortex embodies the same function. Wilson [3] discusses the mapping in the light of additional evidence and examines its potential for pattern recognition. Early related ideas in the pattern recognition literature can be found in Harmon's [4] recognizer and in certain patents [5].

A hypothetical structure (adapted from [3]) schematizing important aspects of the retino-cortical (R-C) mapping is shown in Figures 1 and 2. The "retina" of Figure 1 consists of "data fields" whose size and spacing increase linearly with distance from the center of vision. The "cortex" of Figure 2 is a matrix of identical "message-sending units" (MSUs) each of which receives signals from its own retinal data field, processes the signals, and generates a relatively simple output message that summarizes the overall pattern of light stimulus falling on the data field. The MSU's output message is drawn from a small vocabulary, *i.e.*, the MSU's input-output transform is highly information-reducing and probably spatially nonlinear.

Further, all MSUs are regarded as computing the same transform, except for scale. That is, if two data fields differ in size by a factor of d, and their luminance inputs have the same spatial pattern except for a scale factor of d, then the output messages from the associated MSUs will be identical. (Physiologically, the cortical ***hypercolumns*** [6] are hypothesized in [3] to have the above MSU properties.)

The pattern of connections from retina to cortex is as suggested by the letters in Figures 1 and 2. Data fields along a ray from center to periphery map into a row of MSUs, and simultaneously, each ring of data fields maps into a column of MSUs. The leftmost column corresponds to the innermost ring, the 12 o'clock ray maps into the top row, and so forth.

It is convenient to describe position in retinal space by the complex number $z = re^{i\phi}$, where r and ϕ are polar coordinates. We can denote cortical position by $w = u + iv$, where u is the column index increasing from left to right and v is the row index increasing downwards. For the mapping to have complex logarithmic form, it must be true that the position w of the MSU whose data field is at z satisfies $w = \log z$ or, equivalently, $u = \log r$ and $v = \phi$.

That the equations do hold can be seen from Figure 1. The distance Δr from one data field center to the next is proportional to r itself, which implies that u is logarithmic in r. Similarly, the fact that all rings have equal numbers of data fields directly implies that v is linear in polar angle. Thus (with appropriate units) we have $w = \log z$. (The singularity at $z = 0$ can be handled by changing the function within some small radius of the origin. For present purposes we are interested in the mapping's logarithmic property and will ignore this necessary "fix").

Figures 3-5 (at end of article) review three salient properties of the R-C mapping that have been noted by previous authors. The photos on the left in each figure are "retinal" (TV camera) images. On the right are crude "cortical" images obtained by the expedient of sampling the retinal data field centers. The mapping used has 64 MSUs per ring and per ray.

Figure 3 shows a clown seen at two distances differing by a factor of three. The cortical images, though "distorted", are of constant size and shape. Also shown is the result of rotating the clown through 45 degrees; again, cortical size and shape remain the same. The pictures show how retinal scale change and rotation only alter the *position* of the cortical image. Figure 4 illustrates these effects for a texture. The cortical images are again the same except for a shift. The mapping thus brings about a kind of size and rotation invariance which one would expect to be useful for pattern recognition.

Figure 5, in contrast, shows that the mapping lacks translation invariance. The same clown is seen at a constant distance but in three different positions with respect to the center of vision. Translation non-invariance would appear to be a distinct disadvantage for pattern recognition.

As the clown recedes from the center in Figure 5, its cortical image gets smaller and less defined. The effect illustrates how in a sense the mapping optimizes processing resources through a resolving power which is highest at the center and decreases toward the periphery. This variation is sometimes cited as a useful property of the eye, and was discussed in connection with an artificial retina-like structure by Sandini and Tagliasco [7].

OBJECT-RESOLUTION

The pattern recognition potential of the mapping's size-normalizing property is best seen by defining a somewhat unusual notion of resolution. Recall first that the resolving power ρ of a sensor is the number of distinct signals per unit visual angle; in the case of a linear sensor (such as a TV camera), ρ is a constant. Suppose we ask of a system: when its sensor images a centered object of half-angle A, how many distinct signals, corresponding to the object, will the sensor produce? Let us name this quantity the system's ***object-resolution***, R_o. Then, in the case of a linear system, it is clear that R_o will be proportional to $\rho^2 A^2$. That is, R_o will depend on the distance or "apparent size" of the object, or on the relationship between perceiver and object.

The resulting amount of information may be insufficient for recognition, it may be just right, or it may overload and therefore confuse the recognition process. This uncertainty leads to the scale or "grain" problem noted by Marr [8] and others and to Marr and Hildreth's [9] proposed solution of computations at several resolutions which are later to be combined. The grain problem is also a motivation for the application of relaxation techniques [10] in pattern recognition.

Let us now ask what is the object-resolution of an R-C system. For such a system the resolving power is $\rho = c/r$, with r the distance from the center of vision. The constant c can be defined as the number of MSU outputs per unit visual angle at an eccentricity of $r = 1$. Object-resolution R_o can be found by taking a centered object of half-angle A and integrating over the object from a small inner radius ϵA ($\epsilon \ll 1$) out to A. We have

$$R_o = \int_{\epsilon A}^{A} \frac{c^2}{r^2} 2\pi r dr = 2\pi c^2 ln \frac{A}{\epsilon A} = 2\pi C^2 ln \frac{1}{\epsilon}$$

independent of A.

Thus the mapping's object-resolution or spatial quantization of the seen object is independent of the object's apparent size or distance, and independent of its actual size as well. It depends only on c (and ϵ). Given a fixed value of c, the system may be said to see every centered object, regardless of size, equally well, independent of the perceiver-object relationship. (Strictly speaking, the above integral includes only a fraction $1-\epsilon^2$ of the object, the "outer" fraction. But if ϵ is very small the omitted fraction ϵ^2 will contain an insignificant portion of the object's pattern.)

The object-resolution of the R-C mapping can be thought of in terms of the number of data fields per retinal ring. By mentally superimposing and then expanding and contracting a centered object on Figure 1, one can see that it is examined in an equivalent way at any scale. In fact, it is convenient to use the number of fields per ring as a measure of R_o.

The R-C mapping's *constant* object-resolution is the significant difference between it and a linear system. In the remainder of the paper we will develop implications of this difference. First, why in an important sense the "grain" problem disappears. Second, why Gestalt-like templates are, cortically, suitable for pattern recognition. Third, in outline, how the cortical approach with templates allows a separate adaptive theory due to Holland [11] to be applied to pattern recognition—and in the process may solve the mapping's apparent problem of translation non-invariance.

THE "GRAIN" PROBLEM

Basically, a "grain" problem exists if there is no *a priori* way to tell whether the size of the elements with which the perceiver is looking is the same as that of the optimally informative element of the object or scene. In the linear case, we found that the information about an object may be insufficient, just right, or overloading depending on (1) the perceiver-object relationship and of course on (2) the amount of detail in the object itself.

In the R-C mapping case, the information is *constant*, dependent only on the perceiver. Thus (1) above—uncertainty due to the perceiver-object relationship—disappears. But the information may still, it seems, be insufficient, just right, or overloading—depending on object detail.

We can develop a criterion for the latter as follows. Let an object's "object frequency spectrum" be the two-dimensional Fourier spectrum of a geometrically similar object of unit size, and let f_o be the highest significant (for discrimination) frequency in such a spectrum. Then, roughly, we may say that a mapping with resolution R_o (in units of fields per ring) provides sufficient information about an object if $R_o \geq f_o$.

But this bound is not ultimately limiting. It only says whether information from *one fixation* is sufficient for recognition. Peculiarly, by the mapping's constancy of information, any fixated local *part* of an object is seen in as much detail as is the whole object. Thus, if $R_o < f_o$, the system can always *gather* enough information by scanning, *i.e.*, by moving the center of fixation to any part not seen clearly. R_o is therefore always sufficient, though several fixations may be required.

Can there be too much resolution? Only if objects turn out to be simpler than expected. But often this can be known in advance. In contrast, in the linear case, superfluous resolution will always occur whenever object images become large.

TEMPLATES

In any digital computer implementation, a template for pattern matching consists of a finite (usually rectangular) array of cells in each of which the relative brightness to be matched is specified. The array has a fixed resolution since the number of cells is fixed.

One major traditional problem with templates is a variation of the "grain" problem: Unless the template's resolution is the same as the system's object-resolution, there is virtually no chance of getting a correct match. The R-C mapping offers a solution since the system's object-resolution is fixed, and the resolution of all stored templates can be made exactly commensurate. For instance, the system can acquire its templates by copying its own cortical MSU output images of identified objects. The same objects when later presented in other sizes will be "seen" in the same way.

Templates have other problems, e.g., orientation and brightness variations may lead to mismatch. These will be taken up later. Our analysis suggests, however, that templates may yet have an important role to play in general pattern recognition, provided the matching occurs in a cortex-like space.

OUTLINE OF AN ADAPTIVE CORTICAL PATTERN RECOGNITION SYSTEM

This section will outline a system concept combining the R-C mapping, a production system based on cortical templates, and the theory of adaptation due to Holland.

A visual world mapped as in Figures 1 and 2 suggests a natural polarity between center and periphery. The same centered object, as it grows bigger, expands toward the periphery, and its cortical image, as noted, shifts as a unit from the left side of the "cortex" toward the right side. The implication is strong that processing, in the cortex, should consist of a column-by-column scan [12] from left to right. The pattern of an object, whatever its degree of shift from the left, will be encountered "sooner or later" and thus be available for matching against templates.

Further reflection suggests that rather than working with two-dimensional templates, it might be simpler to use one-dimensional column templates—the identification of a pattern consisting of successive matching of the appropriate column templates. Storage would be saved because a given column template would often be a contributor in more than one two-dimensional match.

An appropriate structure for performing the correlation of successively matching column templates is a form of production system in which (1) the condition of each production includes a column template pattern and one or more internal message patterns, and (2) the action is an internal message to be placed on the common message list. (These internal messages are distinct from the MSU output messages. To avoid confusion, the internal messages will be called i-messages.)

In addition, a separate set of "effector" productions, whose conditions consisted only of i-message patterns, would monitor the i-message list. When an appropriate i-message appeared on the list, the effector would fire. Its "action" would be (1) an external action such as moving the center of vision, or (2) an "internal" action also modifying the system's frame of reference but not directly observable from the outside (more on this later), or (3) a signal to the outside world denoting a pattern name.

Many details need to be filled in to make this an operating system. However, enough has been given to suggest a process in which starting at the left end of the cortex, columns would be scanned and productions would fire in dependent sequence (the dependency based on i-messages as well as the column information being matched), resulting ultimately in an effector firing whose signal named the object in view.

Production systems have not usually been considered in connection with pattern recognition because production conditions typically deal with "normalized" or logical variables and, given the grain problem, patterns in linear vision are anything but normalized. In cortical space, however, patterns *are* normalized so that there the power of productions can potentially be exploited.

But we can go farther. One part of the adaptive theory due to Holland is concerned with "cognitive systems" based on sets of productions called "classifiers". The form of a classifier is, most generally, a string whose condition part consists of a fixed length "environmental detector pattern" together with one or more i-message patterns, and whose action part is an output i-message or effector action. The important point for us is that the "environmental detector pattern" has exactly the form of the column templates we have been considering, so that classifier systems and the adaptive theory may be directly applicable to "cortical" pattern recognition. It has been demonstrated [13-16] that given an appropriate external reward regime a classifier system can evolve a set of classifiers that is adapted to, or "fit", in its environment. This means in particular that the conditions of the classifiers recognize what matters, and the i-messages and actions are appropriate. Much further research must be done, but by combining classifiers with R-C vision, a new path would appear to be open to the objective of a self-organizing visual pattern recognition system.

If the adaptive properties of the Holland system be assumed, we can suggest how the production structure given earlier might deal with non-centered objects. They look different from their centered forms: this is the mapping's translation non-invariance. The problem would be solved if classifiers existed which would react to the off-center form and lead to an effector which would move the center of vision so as to center the object (at which point "standard" classifiers could recognize it).

At first sight, the evolution of this kind of sequence seems implausible: you would need classifiers for every object in every peripheral position. However, the mapping helps by *reducing the detail* seen in an object as it recedes toward the periphery; in the limit, every object becomes just a "blob". This suggests that only a relatively small number of distinct classifiers would be needed to "acquire" any object for standard (centered) inspection.

There remains the problem, not of the isolated object, but of the more-or-less centered one—such as a face—which is still not centered quite well enough to fire its standard classifiers. How can an appropriate centering movement come about? For this question, and related ones, we need to consider the "internal effectors" mentioned earlier.

Three are important in the present discussion: Object-Resolution (OBRES), Azimuth (AZIM), and Brightness Gain (BGAIN). OBRES is an effector (or set of them) which, given appropriate i-messages, will *alter* the system's object-resolution (in effect changing the number of data fields per ring in Figure 1). This permits seeing an object (regardless, of course, of its apparent size) in detail, or more coarsely, depending on the i-message list circumstances. The evolution of OBRES effectors appropriate to different circumstances would occur through the adaptive mechanisms.

If we now recall the problem of the slightly off-center face, it seems plausible that, given some reduced level of object-resolution, most different faces with that degree of decentering could be matched by a relatively small (and thus practical) set of classifiers. These would lead to a movement command bringing the face to the center, where it would be recognized in detail (after, perhaps, restoration by OBRES of a higher R_o).

The AZIM internal effectors set the direction the system regards as "up". In cortical space, this amounts to shifting the input column vector along its length by a definite amount before matching classifier template patterns against it. The purpose of AZIM is, of course, to allow a given set of classifiers to be effective for recognition even if the object is not in standard orientation. But how will the right azimuth be set in such a case? We again have recourse to the evolution of relatively coarse classifiers which, given reduced object-resolution through OBRES, will recognize the presence of a nonspecific ("oblong", say) object at a certain orientation. These would lead to the right AZIM acting, and specific recognition could then occur.

Finally, BGAIN is a set of internal effectors to deal with the persistent problem of setting the right brightness level for template matching. The intent is that the appropriate gain will be determined (via the i-message list) by what is seen, and that the evolution of an appropriate set of BGAIN effectors will again be under adaptive control in the Holland sense.

The various internal effectors, and the external one resulting in movement, are concerned with the system's "point of view" on its visual input, that is, with systematic transformations which will allow the system's form detector set—the classifiers—to function efficiently.

SUMMARY

We began this paper with the retino-cortical mapping and showed how it "saw" centered objects with a resolution independent of the object's size. Constant object-resolution led to a renewed prospect for template matching in general pattern recognition. Fixed size templates permitted the power of production systems to be brought to bear. Finally, the applicability of Holland's adaptive theory to production systems allowed us to suggest that a recognition system based on the mapping might be made self-organizing, in the process overcoming the mapping's "problem" of translation non-invariance.

REFERENCES

[1] Weiman, C.F.R. & Chaikin, G. Logarithmic spiral grids for image processing and display. *Computer Graphics and Image Processing*, **11**, 197-226. 1979.

[2] Schwartz, E.L. Spatial mapping in the primate sensory projection. *Biological Cybernetics*, **25**, 181-194, 1977.

[3] Wilson, S.W. On the retino-cortical mapping. *Int. J. Man-Machine Studies*, **18**, 361-389, 1983.

[4] Harmon, L.D. Line-drawing pattern recognizer. *Electronics*, 39-43, Sept. 2, 1960.

[5] Singer, J.R. Electronic recognition.
U.S. 3,255,437, Jan. 7, 1966.
Burckhardt, C.B., *et al.* Pattern recognition apparatus utilizing complex spatial filtering.
U.S. 3,435,244, March 25, 1969.
McLaughlin, J.A., *et al.* Pattern recognition apparatus and methods invariant to translation, scale change, and rotation.
U.S. 3,614,736, October 19, 1971.

[6] Hubel D.H. & Wiesel, T.N. Uniformity of monkey striate cortex: a parallel relationship between field size, scatter, and magnification factor. *J. Comp. Neurology*, **158**(3), 295-305. 1974.

[7] Sandini, G. & Tagliasco, V. An anthropomorphic retina-like structure for scene analysis. *Computer Graphics and Image Processing*, **14**, 365-372, 1980.

[8] Marr, D. Early processing of visual information.

Philosophical Transactions of the Royal Society of London B, **275**, 483-524, 1976.

[9] Marr, D., & Hildreth, E. Theory of edge detection. *Proc. Royal Society of London B*, **207**, 187-219, 1980.

[10] Davis, L.S. & Rosenfeld, A. Cooperating processes for low-level vision: a survey. *Artificial Intelligence*, **17**, 245-263, 1981.

[11] Holland, J.H. *Adaptation in Natural and Artificial Systems.* Ann Arbor: U. of Michigan Press, 1975.

[12] Evidence and a model for scanning in humans is presented in Wilson, S.W., Strobe imagery: a scanning model. Submitted for publication.

[13] Holland, J.H., & Reitman, J.S. Cognitive system based on adaptive algorithms. In *Pattern-Directed Inference Systems*, Waterman, D.A. & Hayes-Roth, F. (eds.). New York: Academic Press, 1978.

[14] Booker, L. *Intelligent behavior as an adaptation to the task environment.* Ph.D. Dissertation (Computer and Communication Sciences). The University of Michigan, 1982.

[15] Goldberg, D.E. *Computer-aided gas pipeline operation using genetic algorithms and rule learning.* Ph.D. Dissertation (Civil Engineering), The University of Michigan, 1983.

[16] Wilson, S.W. Knowledge growth in an artificial animal. These *Proceedings*.

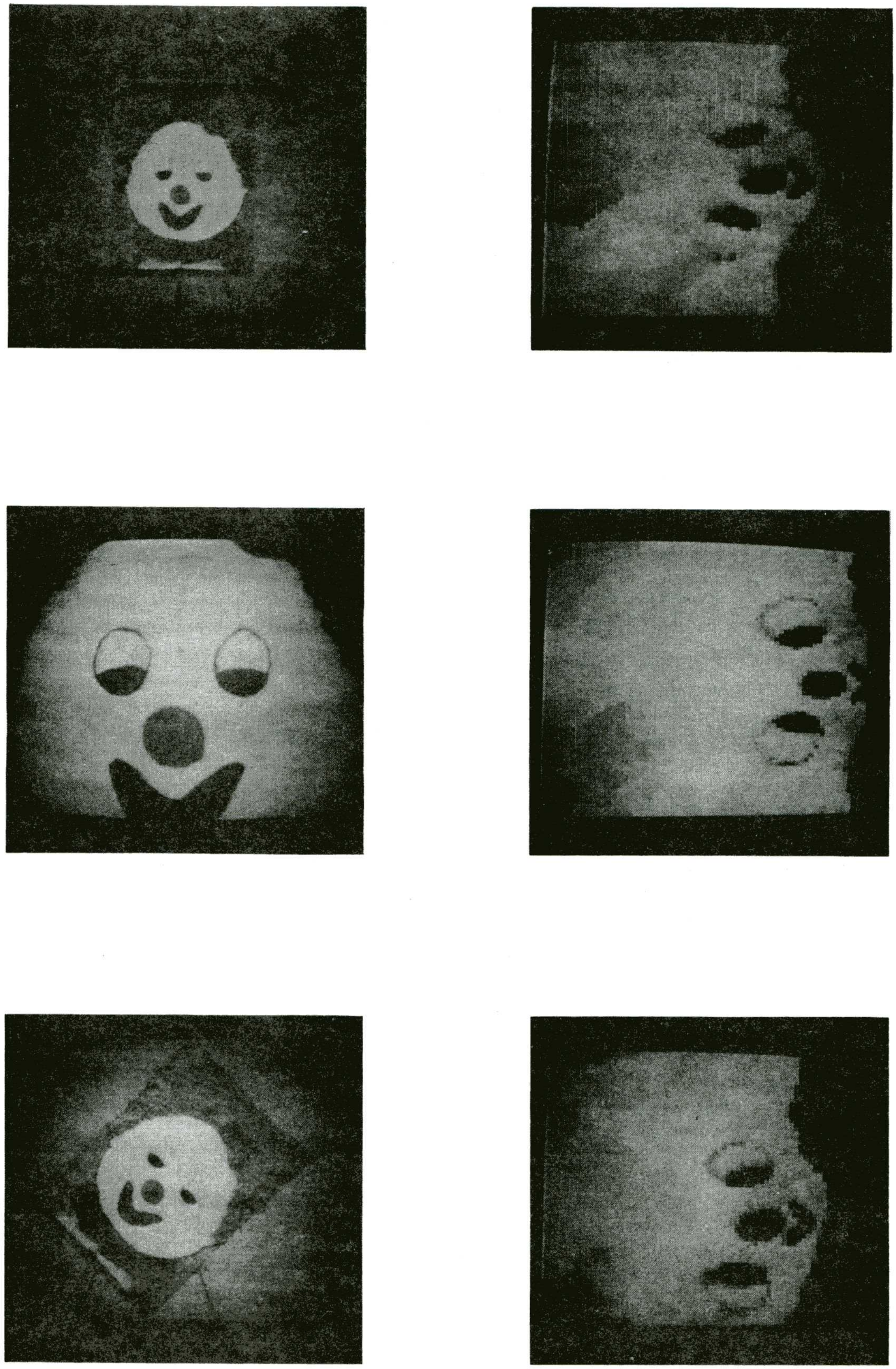

Fig. 3

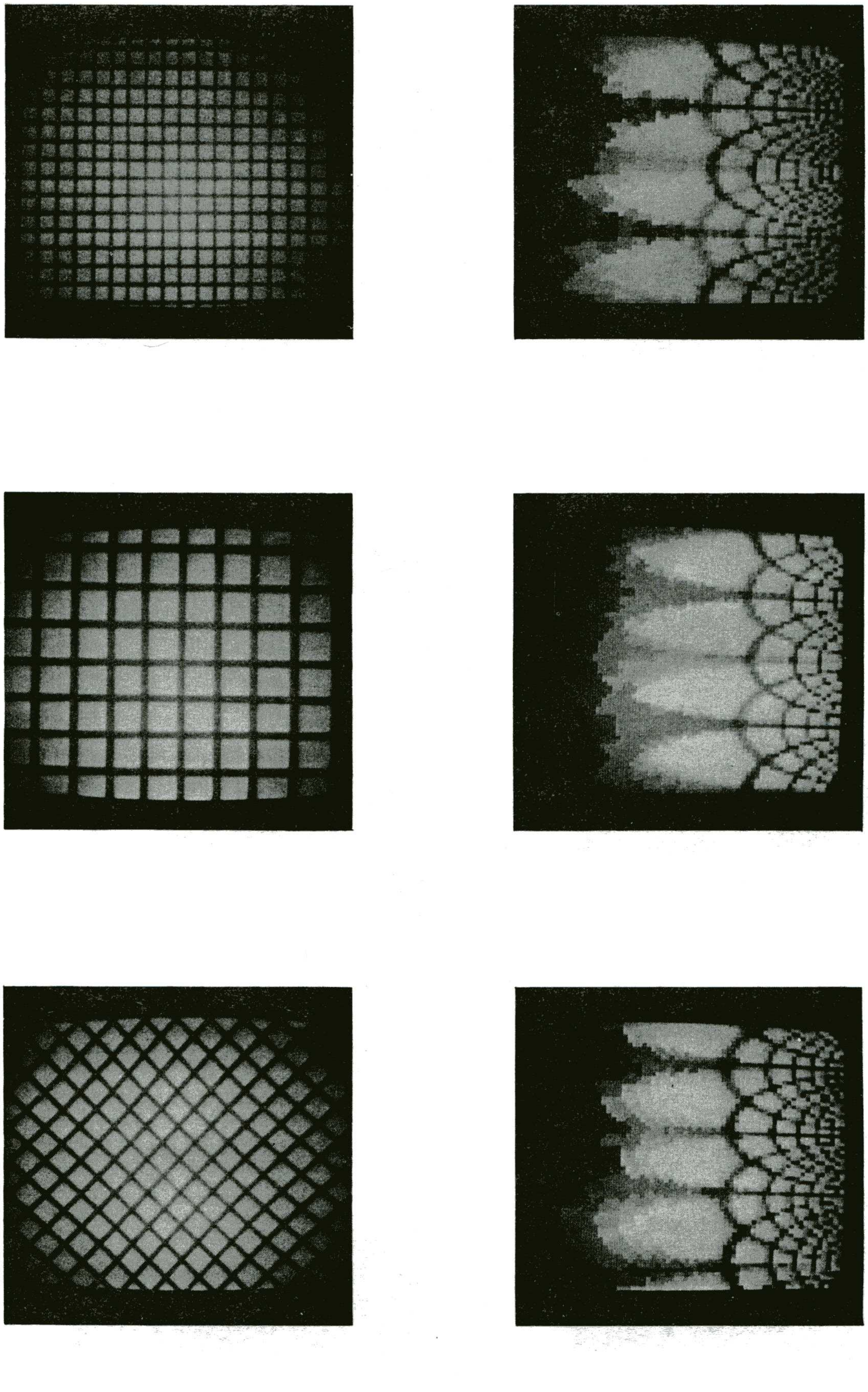

Fig. 4

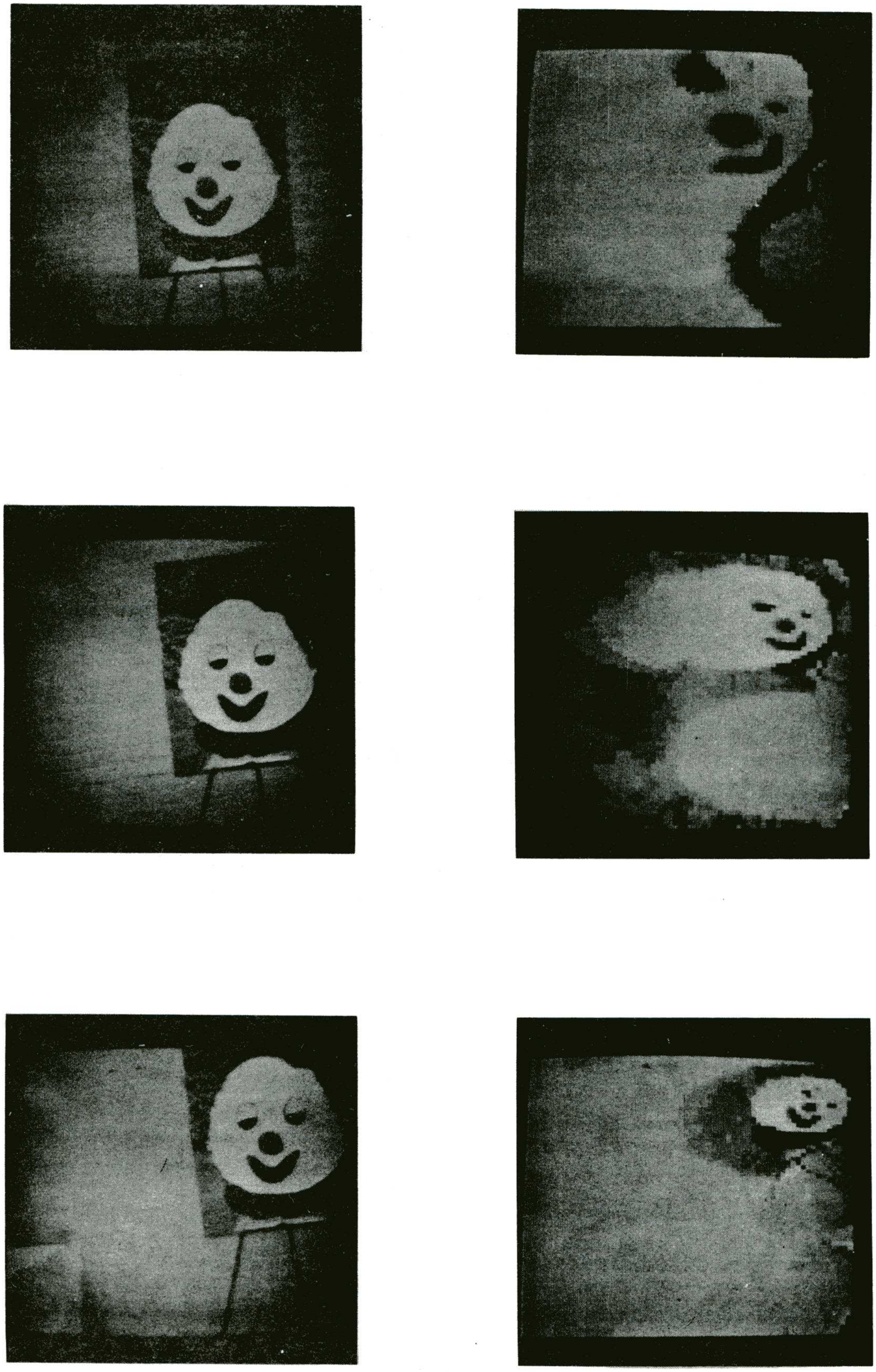

Fig. 5