THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME XXXIX | SEPTEMBER 1960 | NUMBER 5

Binocular Depth Perception of Computer-Generated Patterns

By BELA JULESZ

(Manuscript received March 31, 1960)

The perception of depth involves monocular and binocular depth cues. The latter seem simpler and more suitable for investigation. Particularly important is the problem of finding binocular parallax, which involves matching patterns of the left and right visual fields. Stereo pictures of familiar objects or line drawings preclude the separation of interacting cues, and thus this pattern-matching process is difficult to investigate. More insight into the process can be gained by using unfamiliar picture material devoid of all cues except binocular parallax. To this end, artificial stereo picture pairs were generated on a digital computer. When viewed monocularly, they appear completely random, but if viewed binocularly, certain correlated point domains are seen in depth. By introducing distortions in this material and testing for perception of depth, it is possible to show that pattern-matching of corresponding points of the left and right visual fields can be achieved by first combining the two fields and then searching for patterns in the fused field. By this technique, some interesting properties of this fused binocular field are revealed, and a simple analog model is derived. The interaction between the monocular and binocular fields is also described. A number of stereo images that demonstrate these and other findings are presented.

I. INTRODUCTION

The question of how the two-dimensional projections of the visual world that are supplied to the left and right eyes are matched and combined to reveal the impression of depth is an extremely interesting one. Because of an incorrect analogy derived from measuring distances with a range finder, it is commonly thought that this problem is rather trivial. Admittedly, it is fairly simple to determine binocular parallax by aligning selected portions of an object in the left and right fields of a range finder and computing depth by trigonometrical calculations. The intriguing part of this problem is to explain the remarkable ability of humans to establish *correspondence* between complicated patterns in the two monocular fields. This pattern-matching process is the one being investigated here.

It seems quite clear that patterns perceived in depth afford a promising means for exploring pattern-matching. However, it is well known that the perception of depth under familiar conditions is mediated by many complex cues, both binocular and monocular, which are not easily kept under the control of the experimenter. Thus, many previous explorations have used stereo pictures of familiar objects or line drawings, precluding the separation of interacting cues. The investigation reported here utilized patterns devoid of all cues except binocular parallax, by using artificially created stereo images with known topological properties. Such visual displays ordinarily never occur in real-life situations, and a digital computer (with a video transducer at its output) was programmed to generate them. When these unfamiliar pictures are viewed stereoscopically, peculiar and often unexpected depth effects can be seen. In addition, the perception time of depth under such circumstances is sometimes in the order of minutes (instead of the few milliseconds required for familiar stereo images). This slowing down of the visual process facilitated the present investigation without having much effect on the stability of depth impression after depth was finally perceived.

This paper reports a study of binocular depth perception based upon such presentations. In Section II the problem is posed explicitly and a summary of the results is given. The intent is to provide the essence of this investigation without going into details. The remaining sections are arranged along the sequence of ideas presented in Section II, with the intention of being more specific and of supplying more data. In the last section the new technique of this investigation is evaluated with some possible future applications.

A pair of Fresnel lenses has been enclosed on page 1161 of this issue of the Bell System Technical Journal. They may be used for viewing the stereoscopic illustrations in this paper. Directions for their use may be found in the Appendix.

II. PROBLEM POSING AND SUMMARY OF RESULTS

Human beings exhibit great ability in utilizing binocular parallax to establish the relative depth of objects in the visual field. This process involves finding horizontal shifts between corresponding point domains in the left and right visual fields. The observer seems able to establish this correspondence almost without effort or deliberation, even when the fields differ in brightness and shape (due to reflections and perspective) and in picture material (due to hidden objects seen by only one eye). Thus, depth perception might be likened to the solution of a complicated pattern-recognition problem.

This paper attacks the problem of depth perception as a pattern-recognition problem and poses the following question: In determining binocular parallax do we first recognize monocular patterns in the left and right fields and then fuse them (monocular pattern recognition), or do we first combine the two fields in some manner and then perform all further processings on the fused field [e.g., search for certain patterns (binocular pattern recognition)], or do we utilize a combination of both processes? This question is appropriate both for macropatterns (higher organization of points into objects) and for micropatterns (a few adjacent points). Figs. 1, 2 and 3 attempt merely to illustrate these three possibilities and do not necessarily have relevance to physiological systems.

Artificial stereo images were created by an IBM 704 digital computer. Right and left images were generated, each consisting of 10,000 brightness points, which were assigned one of 16 quantized brightness values at random. In a peripheral "surround" region, the images were identical; in a square-shaped central region, the right-hand image differed from the left by a uniform horizontal displacement. When viewed monocularly, the images appear completely random. But when viewed stereoscopically, this image pair gives the impression of a square markedly in front of (or behind) the surround. By fusing the photographs in Fig. 4 (using two lenses as prisms with a diameter of 2 inches or more and 10 to 18

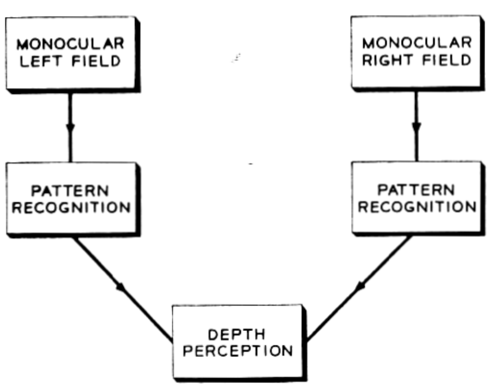

Fig. 1 — Depth perception by monocular pattern recognition.

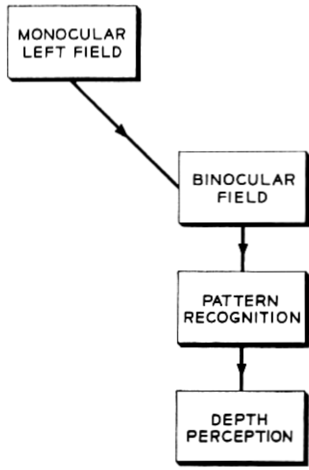

Fig. 2 — Depth perception by binocular pattern recognition.

inches focal length, such as those supplied with this issue) this depth effect can be demonstrated.

Of course, depth perception under these conditions takes longer to establish because of the absence of monocular cues. Still, once depth is perceived, it is quite stable. This experiment shows quite clearly that it is possible to perceive depth without monocular macropatterns. However, if binocular pattern recognition is the principal depth mechanism, the same statement should be true for micropatterns.

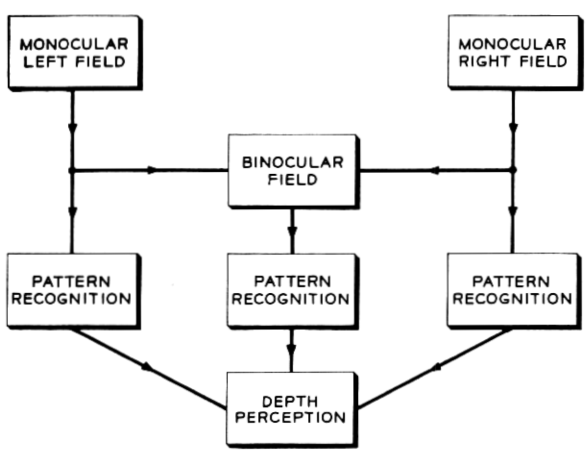

Fig. 3 — Depth perception by monocular and binocular pattern recognition.

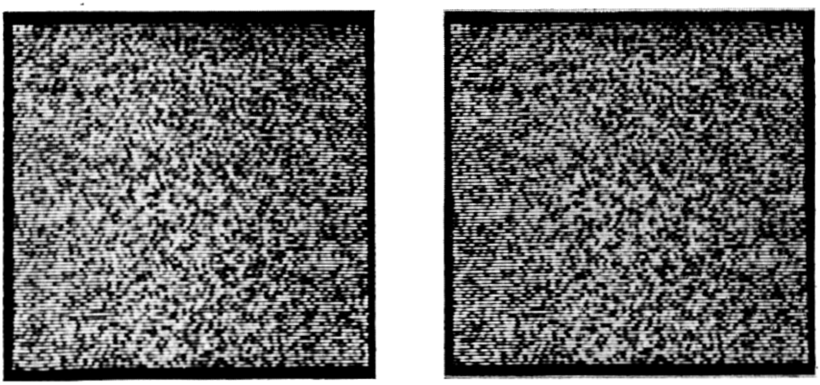

Fig. 4 — Stereo pair with center square above the background.

To study this matter, micropatterns in the stereo pair were drastically altered by blacking out a regular pattern of points in the left field and making the corresponding points white in the right field. Fig. 5 shows the result of this process, where the perturbation grid consists of every even point of every even line. The microstructure of the left and right images is highly different, and yet the center square stands out clearly from the surround.

In spite of the difference in microstructure of the left and right fields, this experiment may not be decisive. It could be argued that the regular perturbation grid is recognized monocularly in its random surround and disregarded, and that the remaining, unaltered points in the two fields possess the same microstructure. It was found, however, that the difficulty of monocularly recognizing the perturbation grid could be increased

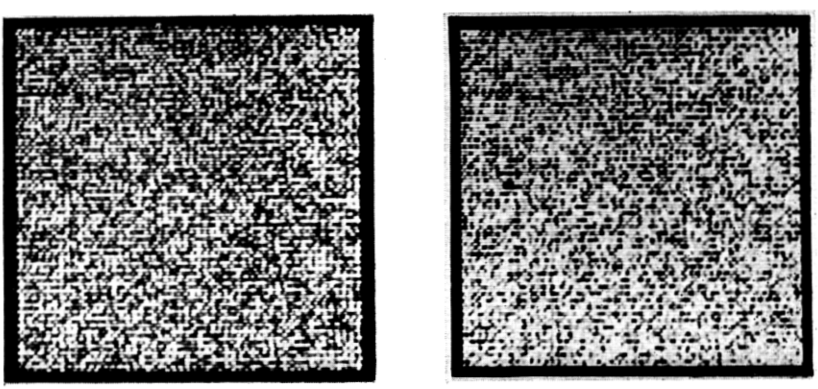

Fig. 5 — Stereo pair with superimposed unmixed perturbation grid.

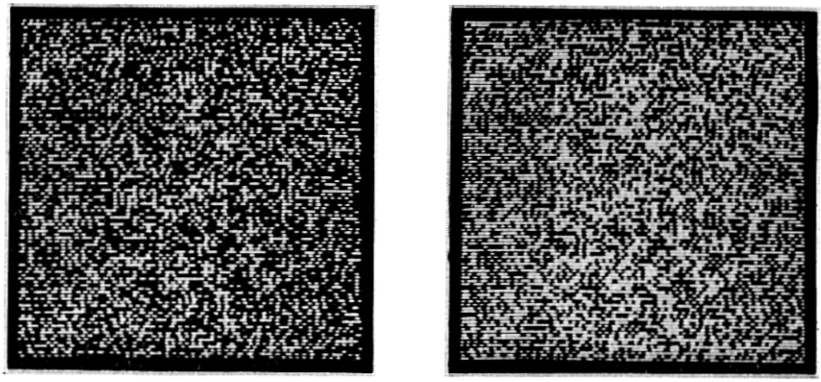

Fig. 6 — Stereo pair as in Fig. 5, but quantized into two levels.

greatly without increasing the difficulty of perceiving depth. For instance, when the random fields are quantized into two levels (black and white), the perturbation grid composed of black (or white) points seems more difficult to find in this surround than in one composed of 16 brightness levels (with many medium grays). The depth effect in Fig. 6 (two-level quantization) can be obtained with the same ease as it can in Fig. 5 (16-level quantization). This makes the assumption of monocularly recognizing the grid very improbable. Together with other evidence (to be discussed in Section VII), it therefore suggests strongly that the two fields are combined first and that the processing is done on the fused field.

Other experiments making use of similar techniques are described. The results shed light on pattern recognition as it is involved in binocular vision. The problem of detecting certain regions in the fused binocular field in order to find depth was particularly investigated. According to these findings, those point domains that are seen in depth (and thus have to be detected in the binocular field) need not possess a *Gestalt*, but the connectivity of the points must be preserved. In the above-described regular perturbation grid, the unaltered points are still connected along one-dimensional arrays (along every other line and column). But if meshlike perturbation grids are applied (which leave the same per cent of points unaltered as in the experiments that will be shown in Figs. 20 and 21, but limit the connectivity of points to small subregions), the depth effect is greatly reduced (as will be seen in Figs. 26 and 27).

As an interesting analogy to certain properties of the binocular field the notion of the difference field is introduced (see Section IX). Although this model is probably very naïve, nevertheless the influence of various perturbations on depth perception often can be predicted by realizing some trivial properties of the corresponding difference field.

The concluding experiments investigate the role of monocular macro-patterns in depth perception. It is shown that their presence greatly enhances the depth effect; thus, monocular and binocular pattern recognition can occur simultaneously as a mixed process. This statement seems to be the final answer to the original problem.

III. BRIEF EVALUATION OF MONOCULAR AND BINOCULAR DEPTH CUES

Depth perception is an interaction of extremely complicated mental processes. These processes utilize certain depth cues which usually are divided into binocular and monocular depth cues. In Table I, a list of these cues is given (without aspiring to completeness).[1]

Most of the monocular depth cues require a tremendous memory capacity; for instance, familiarity with perceived objects implies a catalog of no mean extent.

Binocular depth cues seem simpler and more akin to data processing. Binocular convergence and accommodation are very weak depth cues (as tachistoscopic experiments[2] have shown), and they can be ignored in favor of binocular parallax, which is apparently the principal binocular cue. The invention of the stereoscope[3] strikingly demonstrated man's ability to utilize binocular parallax in order to perceive depth — that is, to determine correspondence between points in the left and right visual fields and measure the horizontal displacements between them. The importance of monocular depth cues in supplementing binocular depth cues is great, as can be demonstrated by the reversed depth effect. It is well

TABLE I

Binocular Depth Cues
Binocular parallax
Convergence of eyes
Correlative accommodation (focusing)
Monocular Depth Cues
Linear perspective (such as converging railroad tracks)
Apparent size of objects of known size (which decreases with distance of observer)
Monocular parallax (change of appearance with change of observer's position)
Shadow patterns (the light-and-shade relations yielding relief)
Interposition (the superposing of near objects on far objects)
Changes due to atmospheric conditions (such as haze, blurring of outlines)
Accommodation (focusing on an object with one eye)
Retinal gradient of texture (decreasing size of texture elements with distance)
Retinal gradient of size of similar objects (rate of decrease of size of houses, fence posts, telegraph poles, etc.)

known that, by interchanging the left and right picture pair in a stereoscope, unfamiliar objects reverse their depth coordinates (far points become near, convex surfaces become concave, etc.). For a familiar object (e.g., a human face) the reversal of depth relationships usually does not take place; that is, the monocular depth cues counteract the binocular ones.

To cancel the effect of these involved monocular depth cues and concentrate on the binocular parallax, most work with stereoscopes uses line drawings for visual stimulus. These drawings comprise simple dots, lines, circles, etc., with different parallax shifts in the right and left fields, and are practically free of monocular depth cues. A vivid depth effect can still be obtained.

The above-mentioned tachistoscopic experiment deserves some additional explanation. A stereo pair consisting of simple line drawings (with parallax shifts in nasal or temporal directions) was flashed for a brief period (in the order of a few milliseconds). Viewing it stereoscopically, subjects could tell almost without any error which of the drawings were in front of or behind a reference plane. This experiment tells nothing about the time required to perceive depth because of the long-persistent afterimages, but it gives some insight into depth processes. First of all, during the short exposure period no convergence or any other motion of the eyes can take place. This fact excludes convergence and accommodation as important depth cues. Second, it demonstrates that during fusion the left and right fields must be labeled, because otherwise the perception of near and far would be confused.

The following investigations are based on the possibility of separating the monocular and binocular depth cues, and concentrate on the problem of how binocular parallax can give the impression of depth.

IV. MACROPATTERN AND MICROPATTERN RECOGNITION; MONOCULAR AND BINOCULAR PATTERN RECOGNITION

It seems clear that a basic aspect of depth perception is recognition of binocular parallax, which consists of a parallax shift between corresponding points in the left and right visual fields. The shift is parallel to the base line (of the eyes); thus, the corresponding points in the left and right fields must lie on the same horizontal line. Now, to determine the exact amount of parallax shift, it is necessary to find the corresponding points in the left and right visual fields. Because the base distance (between the two eyes) and the focal length of the eyes (looking at the stereo pictures at a given distance) are known, there is a simple trigonometric relationship between the parallax shift and the actual depth.

Thus, determining the parallax for every point is analogous to the reconstruction of three-dimensional space. So we come to the kernel of the problem: How can we fuse points in the left and right fields and establish correspondence between them in a stereo sense, when the two fields may differ quite drastically from each other?

The left and right fields of a stereo pair can differ: (a) in brightness (due to different reflections); (b) in perspective (expansion, rotation, shift, etc., of point domains); and (c) by hidden parts (seen only by one eye). Obviously, one is able somehow to find the points in the two fields that belong to considerably different patterns. How is this equivalence established? Do we recognize a face, a square, a few adjacent points, etc., in the left and right visual fields separately and then pick up the corresponding points, or do we first fuse the two fields and perform certain pattern-recognition tasks on this fused field?

To make these questions more precise we introduce the following terminology: Pattern recognition can be divided into two classes. First, *micropattern recognition* concerns simple pattern organizations that take into account some geometrical, topological characteristics in a point's immediate neighborhood. Second, *macropattern recognition* is a higher-order organization of several points. Points grouped together and recognized as a face, square, number, etc., are examples of what is meant by this conception.

The first half of another useful dichotomy is *monocular pattern recognition*, which is performed on the visual field seen by one eye. *Binocular pattern recognition* is performed on the fused field, which is a combination of the left and right monocular fields. It belongs to a special class of processings that incorporate characteristics that intuitively are also important in ordinary (monocular) pattern recognition. Nevertheless, binocular pattern recognition need not necessarily be identical or even similar to monocular pattern recognition.

With these distinctions in mind, we may ask: Is the basic mechanism of binocular fusion a monocular pattern recognition (Fig. 1), or a binocular pattern recognition (Fig. 2), or a combination of both (Fig. 3)? These possibilities multiply when we further differentiate between micropattern and macropattern recognition in each case.

V. DEPTH PERCEPTION WITHOUT MONOCULAR MACROPATTERN RECOGNITION

In aerial reconnaissance it is known that objects camouflaged by a complex background are very difficult to detect monocularly but jump out if viewed stereoscopically. Though the macropattern (hidden object) is *difficult* to see monocularly, it *can* be seen. Therefore, this evidence is not sufficient to prove that depth can be perceived without monocular macropattern recognition.

To investigate this problem, a special visual presentation was created by means of the IBM 704 digital computer and a television transducer developed in the Visual and Acoustics Research Department of Bell Telephone Laboratories.[4,5,6] A pseudo random number routine was programmed to generate random numbers in sequence according to a uniform probability distribution. These numbers were quantized in 16 levels, which were written on tape and then translated by means of a digital-to-analog converter and a special television scanner into 16 brightness levels between black and white. (The quality of present scanning techniques and of photographic processes limits the resolution in brightness, and the final pictures have actually less than 16 identifiable levels.) The television scanner used has the format of a two-dimensional rectangular matrix of 99 rows, each consisting of 105 picture elements. Thus, a picture consists of $105 \times 99 = 10{,}395$ points, whose brightness assume randomly any of the 16 values between the maximum black and white.

A left- and a right-hand stereo image are created by the above-mentioned technique in the following way:

In a peripheral "surround" region, the right- and left-hand images are identical (i.e., the same random brightness points are copied in the two pictures in the same locations); however, in a square-shaped central region, the right image differs from the left by a uniform horizontal displacement. Fig. 7 illustrates this procedure on a small matrix of 6×6 elements. The background points are indicated with small letters having a range of eight letters (brightness values) taken at random. The shifted square in the center has 2×2 elements (indicated by capital

a	b	a	c	d	f
g	e	h	d	c	b
e	f	A	G	a	g
e	a	D	B	e	c
f	c	d	e	f	e
d	g	c	h	b	a

a	b	a	c	d	f
g	e	h	d	c	b
e	A	G	c	a	g
e	D	B	d	e	c
f	c	d	e	f	e
d	g	c	h	b	a

Fig. 7 — Illustration of method by which stereo random pictures are generated.

letters), and the parallax shift in the right field is one picture element to the left.

The distinction between small and capital letters is only for illustration; they possess the same range and distribution, and therefore no macropattern can be seen on any of these random images viewed separately. Those points which are seen only by one eye [e.g., the right side of the square on the right image (c, d)] are generated by the same random number routine.

Fig. 4 showed a stereo pair of 99×105 picture elements, the hidden central square having 40×40 elements, and the parallax shift (Δ) being four picture elements. Both of these pictures, viewed separately, give an entirely random impression, and only an experiment can determine whether when fused stereoscopically the center square will be seen in depth in front of (or behind) the surround.

The images presented can be fused easily by using two simple lenses (of more than two-inch diameter and 10- to 18-inch focal length) as prisms. After fusion, there is a vivid depth effect. The square is in front of the background plane, and the depth impression is very stable. It is interesting that the depth effect does not appear at once, but appears only after a fairly long time in comparison to that in familiar stereo pictures. A curious learning process can be experienced; that is, the time required to get the depth effect diminishes after repetitive trials. The problem of what is really learned here is an interesting question in itself and deserves further investigation.

The fuzziness of the edges of the square is mainly due to the fact that, by chance, some of the brightness points along the edges of the square can belong to both the square and the background, and there is a tendency to interpret them ambiguously. The probability that two or more adjacent points should become ambiguous is very low, so the fuzziness of the edges is about ± 1 picture element in width. (These rough edges reveal that no "*Gestalt* organization" takes place in binocular fusion though the square has a "good *Gestalt*.")

Fig. 8 demonstrates another stereo pair generated in the above way by the computer, but now there are three planes: the background plane, a central rectangle 60×40 in size and with a parallax shift of $\Delta_1 = 4$, and a third rectangle 20×40 in size and with a parallax shift of $\Delta_2 = 8$. It takes some time to get the bigger rectangle in front of the background, but it usually takes even more time to get the smaller rectangle in front of the bigger one. After the three different planes of depth are perceived they remain very stable. The same is true for the reversed depth effect. If the left and right images are interchanged (thus the parallax shifts

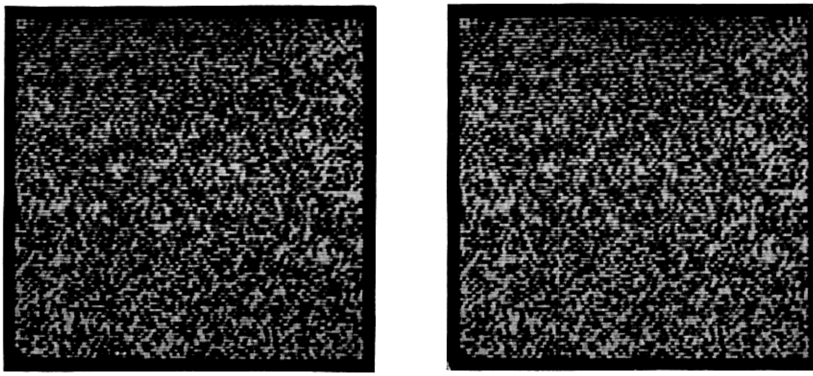

Fig. 8 — Stereo pair with two different planes of depth above background.

are not in the nasal direction but in the temporal one), the three planes reverse their depth relation to the observer. If highlights are eliminated and the surface-like appearance of the pictures is reduced, no monocular depth cues remain, and the reversed depth effect can be obtained with the same ease as in the regular case (Fig. 9).

Apparently, the greater difficulty in seeing the smaller rectangle at its "proper" depth arises, not because of its greater parallax shift, but merely because of its smaller size. By using the same parallax shifts as in Fig. 8 but increasing the size of the closest rectangle and decreasing the intermediate one, it can be demonstrated that the closest rectangle emerges first from the background followed by two smaller ones behind on the sides (see Fig. 10).

These experiments show that it is possible to perceive depth without

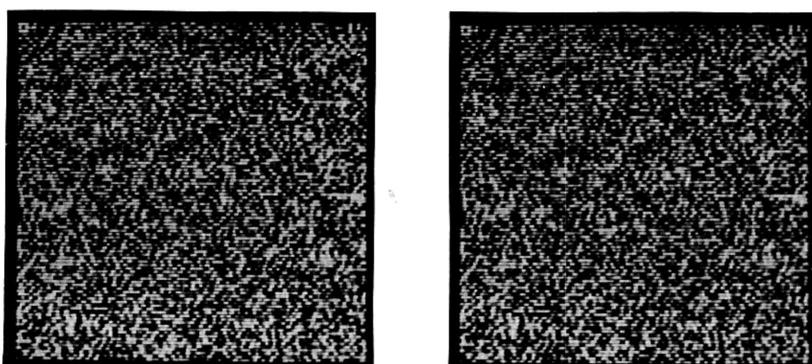

Fig. 9 — Stereo pair with two different planes of depth behind foreground.

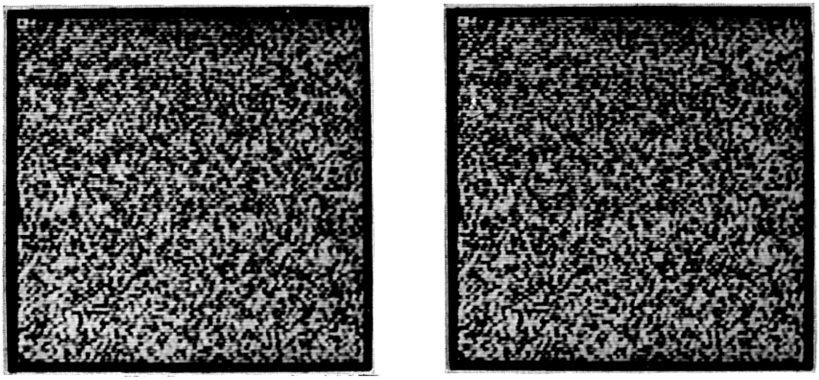

Fig. 10 — Stereo pair with two different planes of depth above background.

monocular macropattern recognition. We must now investigate this same matter for micropattern recognition, if the flow chart for depth perception is to be established. In Sections VI and VII this problem is investigated.

VI. EFFECTS OF INTRODUCED PERTURBATIONS ON THE DEPTH PERCEPTION OF STEREO RANDOM FIELDS

If we compare ordinary stereo photographs of real-life objects, the left and right pictures can differ substantially without being difficult to fuse.

In the present investigation we concentrate only on local perturbations, such as differences in brightness, and ignore the problem of differences in perspective (expansions, rotations, etc.), which belongs to the class of perturbations extending over the pictures according to complicated laws.

The perturbations were introduced in only one of the two pictures, leaving the other unchanged. The perturbations naturally have an effect on the general appearance of the fused image and on the stability of depth perception, but these are not really the effects we are interested in. Our basic question was to find out whether or not, after a given type and amount of perturbation, depth could still be perceived. In other words, to what extent can the brain solve the problem of pattern-matching after distortions are introduced?

In the following investigations some limitations are imposed on the input material. The random stereo images contain only point domains with a uniform parallax shift. The value of the parallax shift and the

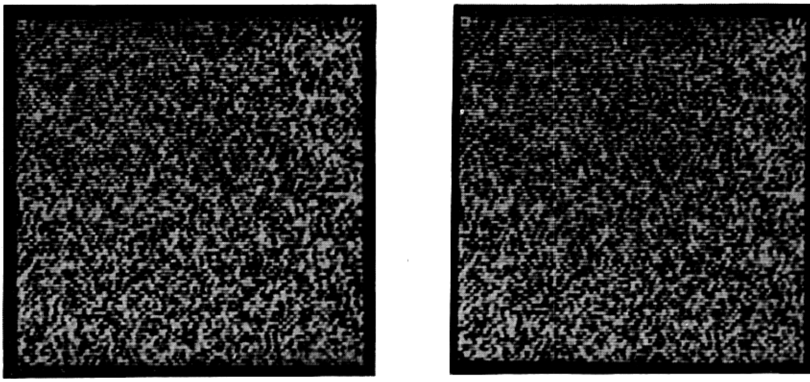

Fig. 11 — Stereo pair with gaussian noise perturbation (14-db signal to noise).

size of the center square is kept constant. The stereo pair before perturbation is like Fig. 4, i.e., two 40×40 squares with $\Delta = 4$.

The first type of perturbation introduced was the addition of gaussian noise on one of the stereo images. In Fig. 11 gaussian noise is added to the left picture. The signal-to-noise ratio (peak-to-peak signal to average noise) is 14 db. Nevertheless, the square is clearly visible in depth though several ambiguous points on the background and the square give rise to a lacy appearance. Even with a perturbation of 6 db signal to noise, the depth effect can be obtained, although the image is markedly deteriorated. Some additional findings will be discussed in Section IX.

Another type of noise is introduced by quantizing one of the stereo pairs in fewer levels than the other image. In Fig. 12 the left picture is

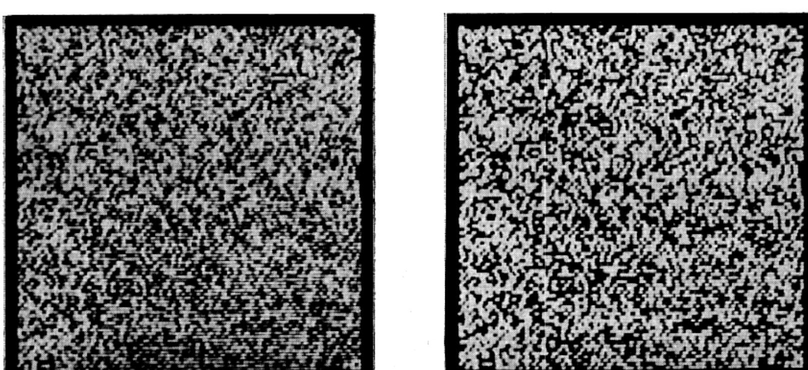

Fig. 12 — Stereo pair with quantizing noise perturbation.

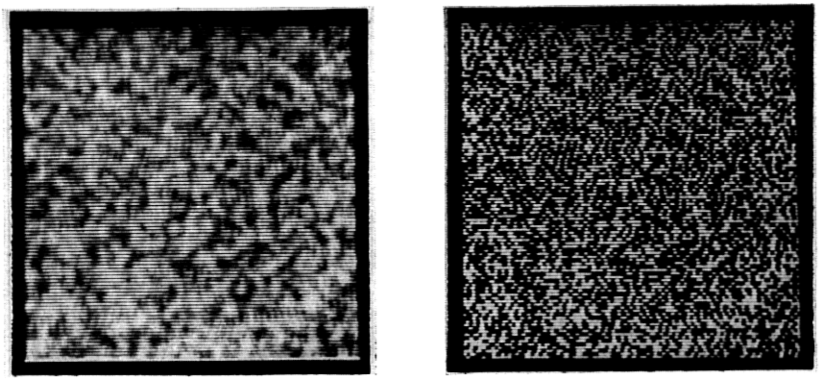

Fig. 13 — Stereo pair with blurred left picture.

quantized only into two levels (black and white). A decision level in the middle gray was chosen, and whenever a brightness point was greater than this it was represented as white, otherwise as black. The right picture is not altered and has 16 brightness levels (actually, on the photograph reproduced here, it has less, but there are more than four). This perturbation, in effect, yields to a special type of noise, sometimes called quantizing noise, and by fusing the stereo pair of Fig. 12 it becomes apparent that even this disturbance does not cancel the depth effect.

The next experiment uses a random stereo pair similar to Fig. 4 (but both the left and right images are quantized into two levels), and the left image is blurred (see Fig. 13). The blur is introduced in the computer by taking each point of the original image and adding to it its surrounding points with equal weights. The blurred u_i^* brightness points of Fig. 13 were obtained according to the following operation:

$$u_i^* = \frac{1}{9} \sum_{j=0}^{8} u_{ij},$$

using the notations in Fig. 14.

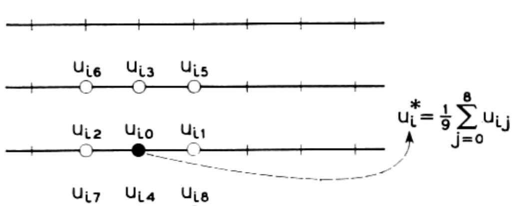

Fig. 14 — Illustration of the method by which blurring was introduced.

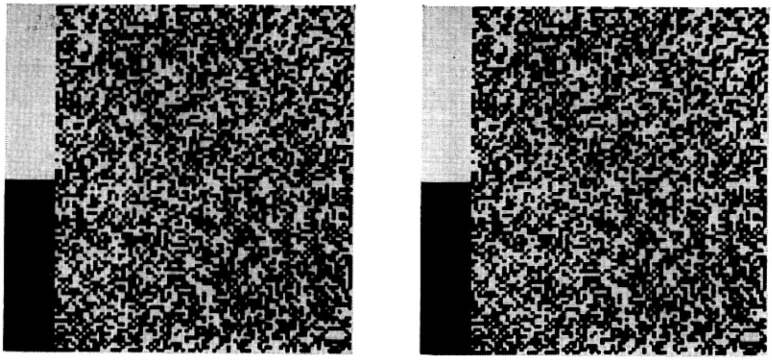

Fig. 15 — Stereo pair with positive pictures quantized into two levels.

This amount of blurring reduces the information content of the left image considerably, but it is still enough to carry the depth information. What is more, the eye is able somehow to see the whole as a sharp picture.

The following experiment is instructive in itself, and will be referred to in the next section. Fig. 15 shows a stereo pair (as does Fig. 4), but both left and right pictures are quantized into two levels. Depth can be easily perceived. Now in Fig. 16 the left picture is identical to the left picture in Fig. 15, but the right picture is the negative of the right picture in Fig. 15. Thus, all points are complemented. Experimenting with Fig. 16, we can conclude that it is not possible to fuse a positive and a negative picture. In addition, strong binocular rivalry can be experienced.

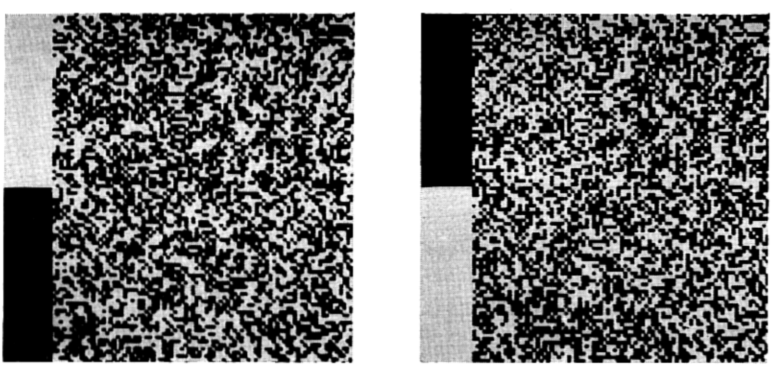

Fig. 16 — Stereo pair with positive and negative pictures quantized into two levels.

In these presentations, special care was taken to ensure uniformness of the black and white values. (To avoid filter ringing, we used a "sample and hold" circuit without filter in the digital-to-analog converter). The bars on the left side illustrate the effect of fusion and rivalry of more extended uniform areas. This experiment shows that one of the greatest perturbations we can introduce is to use maximum black or white points in one field and their complements in the other.

VII. DEPTH PERCEPTION WITHOUT MONOCULAR MICROPATTERN RECOGNITION

The perturbations introduced in the previous section were not drastic, and so the corresponding micropatterns in the left and right images still had some resemblance to each other. Nevertheless, it is apparent that fusion is not the result of a simple point-to-point correspondence between the stereo images. At least, certain coding operations that enhance the resemblances between corresponding micropatterns are required before fusion.

In the next experiments, the resemblance between the left and right micropatterns is drastically reduced; despite this fact, depth can be perceived in several situations.

In all the experiments that follow, the original stereo image is identical to the one in Fig. 4, with either 16 or two brightness levels and $\Delta = 4$. Then, a regular grid is superimposed on the left and right random fields, as shown in Fig. 17.

Every second point in every second line (shaded squares) is changed to maximum black in the left field and to maximum white in the right field. As shown, 25 per cent of the points are so treated, with the result that these points cannot be fused. The rest is unaltered. This arrangement of the perturbation grid removes similarities between the micropatterns of the stereo pairs in the following sense: There are not any corresponding points in the left and right images which have an identical neighborhood. At least one point is changed to its complement in any micropattern 2×2 or greater in size. Fig. 5 shows such a stereo pair of 16 brightness levels having a black grid in the left field and a corresponding white grid in the right field.

The grid cannot be seen monocularly, since it is embodied in the random field. When Fig. 5 is viewed binocularly, however, the square jumps out and is quite stable.

This experiment is still not decisive. One might argue that the resemblance between corresponding micropatterns is not completely removed because, along every other horizontal or vertical line (these are

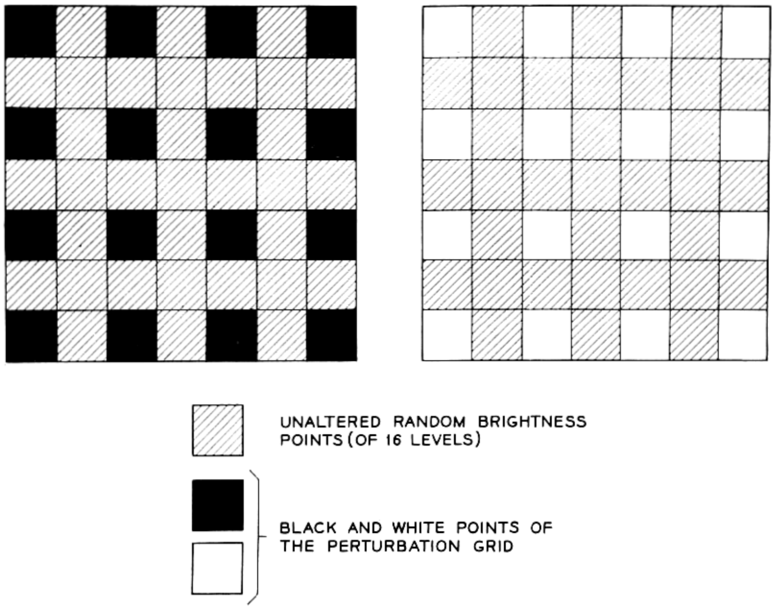

Fig. 17 — Illustration of the method by which the unmixed regular perturbation grid was generated.

unaltered), the micropatterns in the two fields are identical. A searching operation might exist that finds in the left and right monocular fields such identical one-dimensional arrays. To investigate this objection the following experiment was performed:

The same regular perturbation grid of Fig. 17 was used, but with a modification. Instead of uniformly blackening out all of the grid points in the left field, these points were made black or white at random. Then, the corresponding points in the right-hand field were assigned the complementary values (see Fig. 18).

Fig. 19 shows a 16-level random stereo field with this kind of mixed regular grid. Under these conditions depth is *not* perceived. Because in both perturbations (according to Fig. 5 and Fig. 19) the same points are left unaltered in the left and right fields and the same points are also perturbed, the fact that depth can be perceived in one case and not in the other removes the above objection.

Even this experiment is not a final proof that monocular micropattern recognition does not play some part in fusion. It might still be argued that this striking difference between depth perception, using for perturbation the unmixed grid (Fig. 5) or the mixed grid (Fig. 19), could be

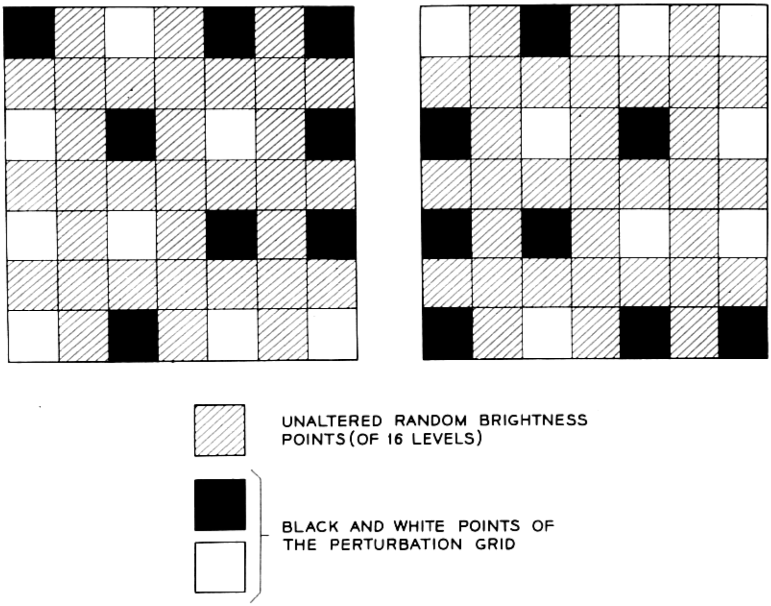

Fig. 18 — Illustration of the method by which the mixed regular perturbation grid was generated.

explained by this hypothesis of monocular pattern recognition: In the unmixed case the regular grid might be recognized monocularly by an unconscious process, then disregarded, and the remaining random points could now be fused monocularly without any difficulty. In the case of the mixed grid, this grid is not apparent monocularly, so the removal of

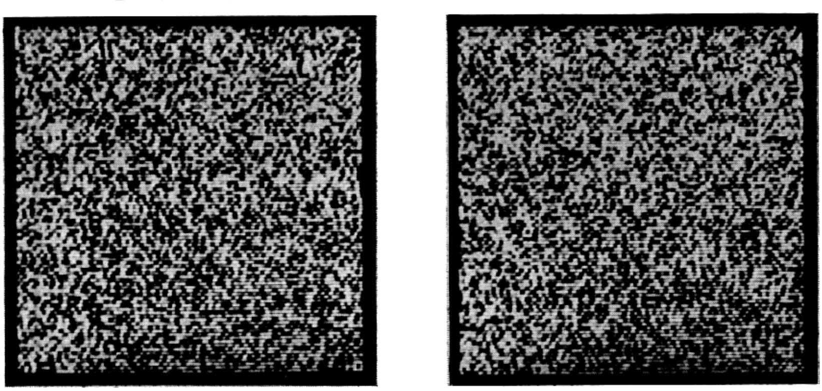

Fig. 19 — Stereo pair with mixed regular perturbation grid.

the grid points is not possible, and no fusion can take place. This hypothesis seems very improbable. Even to suppose that the regular grid can be recognized and removed unconsciously is unlikely, but, in addition, the monocular recognition of certain regularities in random fields would require extremely complex operations (e.g., autocorrelation technique detects only the periodicities of the hidden regularities without determining the location of the grid points). Even assuming that such a process exists, it certainly could find a regular grid composed of maximum black (or white) points much more easily in a surround of random brightness points of 16 levels (with many medium grays) than it could in a surround having only black and white random points. To check this assumption, we used the unmixed regular grid of Fig. 5 with only the modification of quantizing the random fields into two levels. Fig. 6 shows this case, with the result that depth can be perceived even sooner than with 16-level quantization, which disproves the assumption of monocular recognition and removal of the regular perturbation grid.

The stereo pair in Fig. 20 originally had a random field quantized into two levels, and a checkerboard-like perturbation grid was superimposed as illustrated in Fig. 21. Here, 50 per cent of the total points are complemented, and the regular grid has a double periodicity. Even in this case the depth effect can be easily obtained by fusing Fig. 20.

In these last experiments, the left and right images differ from each other considerably and the monocular recognition of the perturbation grid is made very difficult, yet we can still fuse the unaltered points with ease. These results disprove the hypothesis of monocular pattern recognition (both in the micro and macro sense), and suggest the second alternative: that the two fields are first combined and all further processings are performed on the fused binocular field.

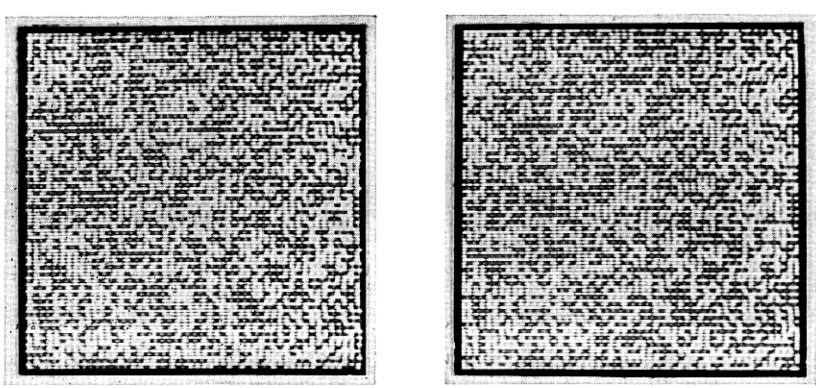

Fig. 20 — Stereo pair with "checkerboard" perturbation grid.

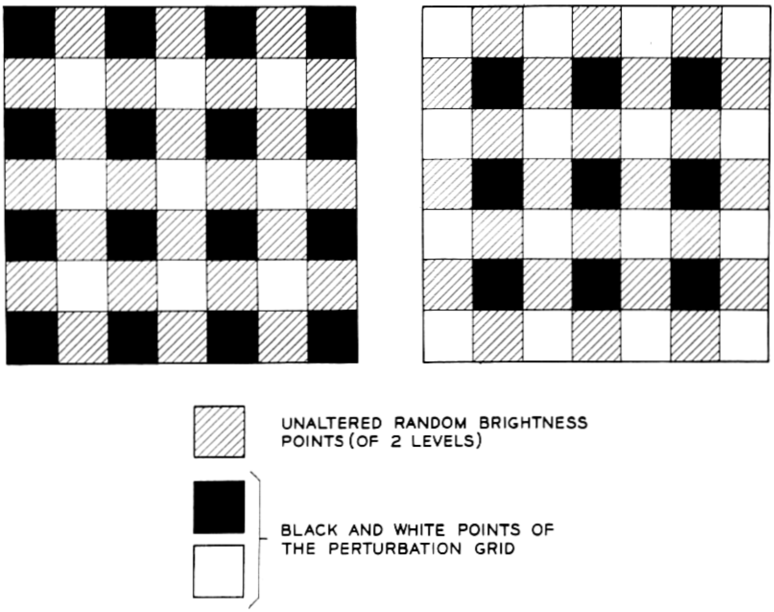

Fig. 21 — Illustration of the method by which the "checkerboard" perturbation grid was generated.

VIII. THE CONNECTION BETWEEN BINOCULAR PATTERN RECOGNITION AND DEPTH PERCEPTION

The demonstrations in the previous sections strongly suggest that, under these conditions, the perception of depth utilizes certain processings performed entirely on the fused binocular field. We intentionally do not yet call these processings binocular pattern recognition, because we must first investigate the feasibility of some processes that in ordinary usage are regarded as simpler than pattern recognition.

It has already been shown that matching corresponding point domains in the two fields does not require organizing these point domains into a higher entity of monocular macropatterns or micropatterns. One might think that the matching of corresponding point domains (instead of corresponding patterns) could be achieved by searching for a best fit according to some similarity criterion (e.g., maximum cross-correlation). A simple way to find correspondence between points in the two fields is to select a zone (of arbitrary shape) around any point in the left field and search for a zone in the right field (having the same shape) that is most similar to the left zone according to a given criterion. If this zone-matching were performed for every point in the visual field and each point were assigned the parallax shift (or depth value) so obtained, the final three-dimensional representation could be achieved. But such a process cannot work. If the zone size is small, noise can easily destroy any zone-matching; if the zone size is increased, ambiguities arise at the boundaries of objects which are at different distances. For instance, this process could never detect a one-dimensional array in front of a background plane, which is relatively an easy task for a human.

A more sophisticated version of this processing would be to vary the shapes of the zones during the zone-matching; finding a best fit would determine both the corresponding zones and their shapes. Now, in the absence of monocular cues, to search for a best fit and simultaneously vary the shapes in all possible ways seems a very inefficient and time-consuming operation. In addition, some of our previous results make such processes seem less than likely. For instance, in the case of the unmixed perturbation grid (Fig. 17) — where depth was perceived — we could imagine that a zone having the shape of a horizontal (or vertical) array might be found. But the same process would also have selected the same zone shape and properly matched these zones in the case of the mixed grid (Fig. 18), although depth was not perceived in this case.

Thus, it seems difficult to find simple operations (avoiding the use of pattern recognition) that give depth information consistent with that abstracted by the human visual mechanism. However, it is possible to demonstrate certain properties of point domains that are necessary in order for them to be seen in depth. These properties incorporate concepts such as connectivity, minimum size of a point domain, organization of close or periodic parts in higher entities, etc. We intuitively associate these notions with pattern-recognition operations. Therefore, our findings suggest that, under certain conditions, the perception of depth depends upon binocular pattern recognition. There is, of course, no evidence that this pattern recognition on the binocular field is identical to ordinary (monocular) pattern recognition. Nevertheless, an understanding of binocular pattern recognition may well be revealing when the broader aspects of pattern perception are considered. We will proceed, therefore, to investigate certain properties of patterns in the binocular field that yield depth effects.

The first question usually raised is this: Must the point domains possess any familiar pattern for them to be seen in depth? The answer is no. Any connected point domain can be seen in depth regardless of the shape of its boundary. The point domain should be connected at least in one dimension. This one-dimensional connectivity is a trivial property, which every object in real life possesses, and the following experiments show

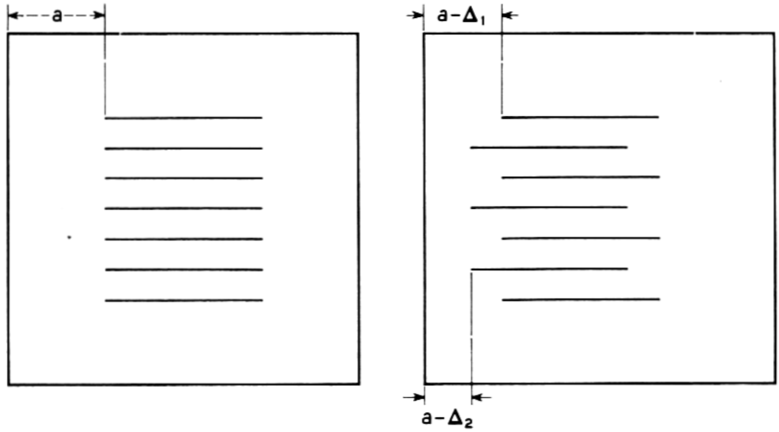

Fig. 22 — Illustration of the method by which a transparent center square was generated above another square (using horizontal arrays).

that this important property is preserved in the binocular field. Fig. 22 demonstrates the way in which a random stereo field (Fig. 23) is generated, with every even line (of 40-picture-element length) having a parallax shift of $\Delta_1 = 4$, and every odd line having one of $\Delta_2 = 6$.

The even and the odd lines each form a square that can be seen in depth; the far one appears to have a regular surface; the closer square seems transparent. Either horizontal or vertical connectivity yields the same results. Fig. 24 shows such a case, where the pattern is composed of vertical random arrays of 40 picture elements in length. Twenty even

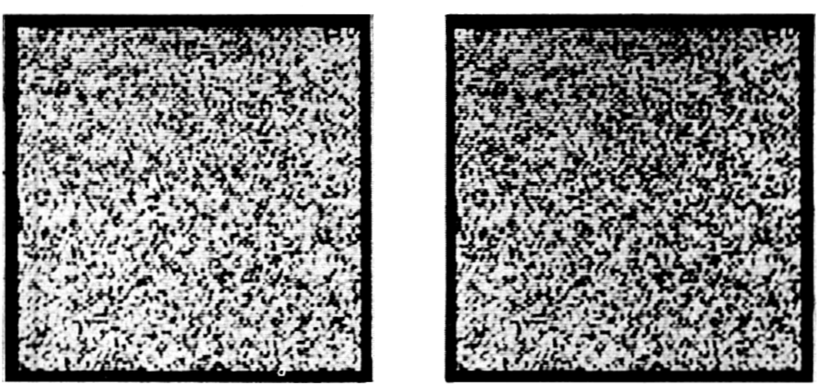

Fig. 23 — Stereo pair with a transparent square (composed of horizontal arrays) above the center square.

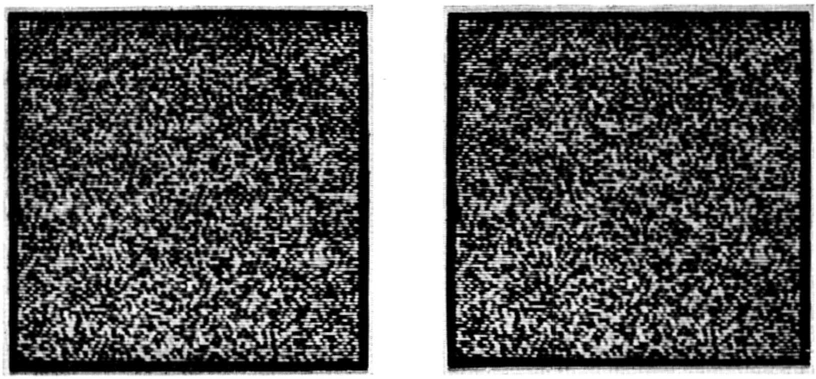

Fig. 24 — Stereo pair with a transparent square (composed of vertical arrays) above background.

vertical arrays form the "transparent" center square; the odd vertical arrays belong to the background.

If we now try an experiment using isolated points of the same depth, it is very difficult to see these points forming a ghost-like plane, even if the points are regularly spaced. Fig. 25 shows such a case, where the regular presentation of Fig. 4 is used but every second point in every second line has a parallax of $\Delta_2 = 2$. If these isolated points at the same distance are not regularly spaced and not dense enough, they cannot be organized as forming one surface.

To show the importance of connectivity in another example, Fig. 26 demonstrates a stereo pair with a meshlike perturbation grid (shown in Fig. 27). Although 50 per cent of the points are unaltered (as in Fig. 20),

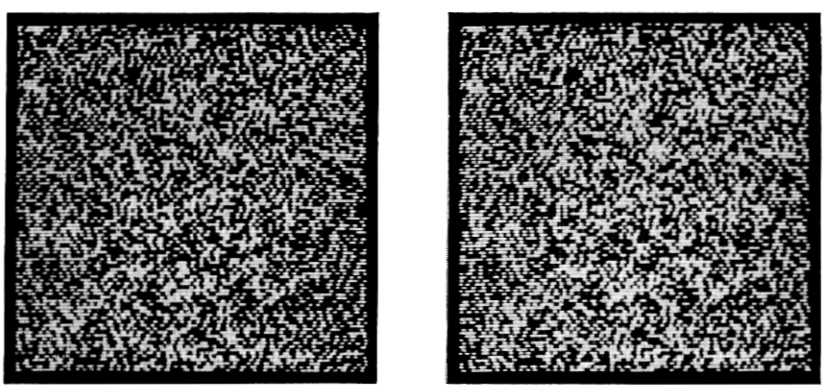

Fig. 25 — Stereo pair with "ghost" square (composed of isolated points) above background.

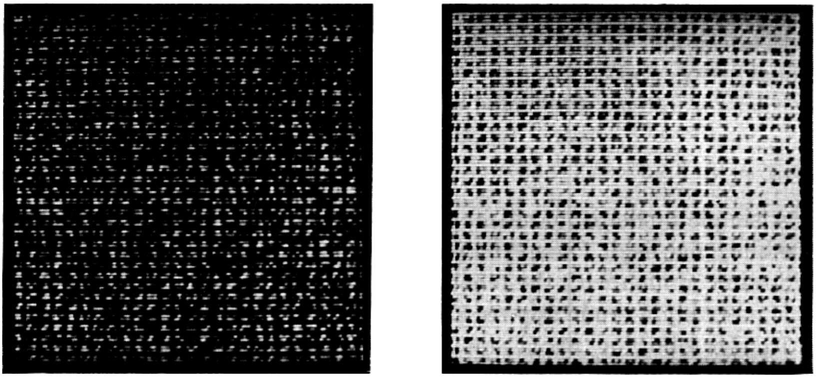

Fig. 26 — Stereo pair with meshlike perturbation grid.

the depth effect is now greatly reduced. The only explanation offhand is the fact that the perturbation mesh limits the connectivity to small, separated subdomains. It is also interesting that these subdomains must possess a critical size in order to be seen in depth. The investigation of this quantitative aspect is not attempted at the present.

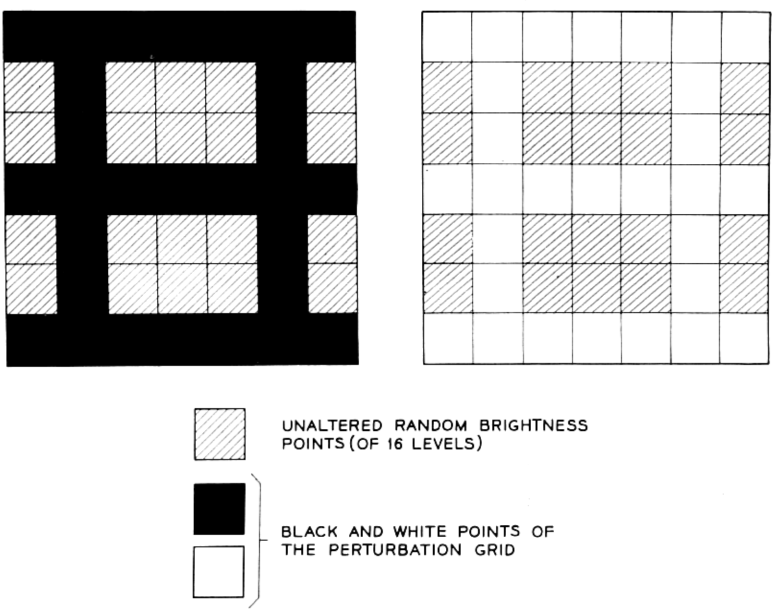

Fig. 27 — Illustration of the method by which a meshlike perturbation grid was generated.

These findings might suggest that the patterns seen in the binocular field are similar to contour lines, which consist of continuous one-dimensional arrays and connect the points of equal parallax shift. In the next section a simple analog model will be derived along these lines.

IX. THE DIFFERENCE FIELD AS A SIMPLE ANALOG TO THE BINOCULAR FIELD

A simple model is an aid in getting greater insight into properties of the binocular field. The model that follows appears to have several properties in common with the binocular field as perceived, but on the whole it is probably a crude approximation.

In the following we accept the assumption that binocular pattern recognition is performed entirely on the binocular field in order to derive depth information, and we remember that the image points belonging to the left and right fields must be labeled. The binocular field $f(L, R)$ is a function of the left and right fields (L and R); thus, the set of all points in the binocular field is a function of the set of brightness points $L(x, y)$ and $R(x, y)$ in the monocular fields, where x and y are the coordinates.

Now the value of $f(L, R)$ at some point x, y must not be merely a function of $L(x, y)$ and $R(x, y)$, but must in fact depend on the values of L and R at other points. Thus, it must not be of the form

$$f(x, y) = f[L(x, y), R(x, y)]$$

because the crucial information, namely, to which field a certain point originally belonged, would be lost thereby.

In the previous section it was shown that cross-correlation also cannot be the combining operation between L and R. To derive a simple model of the binocular field, we generalize the notion of cross-correlation, with L and R being combined in the following way:

$$g_k(x, y) = L(x, y) * R(x + k, y),$$

where k is a positive number referring to a given horizontal shift to the right and $*$ refers to an operation (as yet unspecified). We call the set of all g_k functions (as k varies in a given range) the *analog binocular field* and all further processings will be performed on this field. We call this processing *binocular pattern recognition* without further specifying it at the present.

To be more specific, we now choose a particular $L * R$ by demanding that it be a simple operation. Addition or multiplication seems less favorable than substraction or division; this assumption is based on the experiments with Fig. 17, where the perturbation with an unmixed grid gave depth effect, and Fig. 18, where the mixed regular grid did not. Neither $g_0 = L(x, y) + R(x, y)$ nor $g_0 = L(x, y) \cdot R(x, y)$ would discriminate between Fig. 17 and Fig. 18 (being identical for both cases), whereas both $L(x, y) - R(x, y)$ and $L(x, y)/R(x, y)$ could account for the difference in depth impression.

Finally, we choose $D_k(x, y) = L(x, y) - R(x + k, y)$ as the simplest operation at hand, and call D_k the difference field having a parallax shift of k picture elements. The set of all D_k fields is an analog binocular field, which is designated as the *difference field* D. In these investigations, we limit k to integers in a given range; thus, the final model consists of a finite number of difference fields of different parallax shifts. Now, determining the binocular parallax is equivalent to finding patterns in some of the D_k fields. We called this processing binocular pattern recognition, and in this analogy we regard it as being identical to *ordinary (monocular) pattern recognition.*

In the case of our regular presentation (that is, the random stereo field containing a square with a parallax shift of four picture elements surrounded by a background with zero parallax shift), the following difference fields will be obtained: (a) D_k for $k \neq 0$ or 4 are random fields where each brightness point has a triangular probability distribution [this is the result of taking the convolution between the two uniformly distributed random variables L and R, which gives the triangular probability distribution of $(L - R)$]; (b) D_0 will be zero for every point in the background and will be random for every other point, that is, for the square and for points seen only by one eye; (c) D_4 will be zero for the central square and random elsewhere. (D_0 and D_4 are shown as the left and right pictures in Fig. 28.) Here the zero difference corresponds

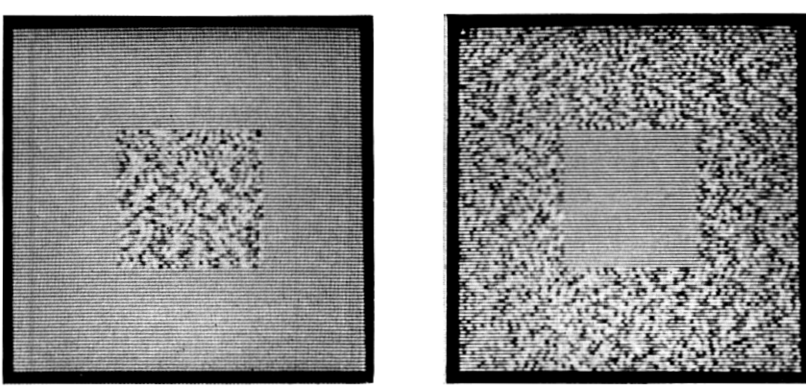

Fig. 28 — Difference fields D_0 and D_4 for the case of Fig. 4.

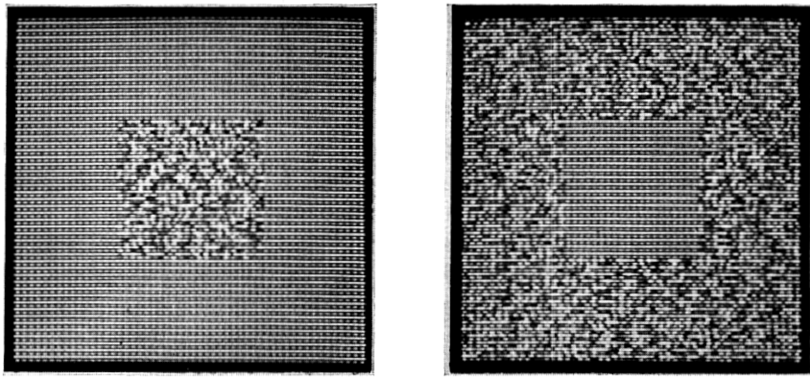

Fig. 29 — Difference fields D_0 and D_4 for the case of Fig. 5.

to a medium-gray level, the maximum positive difference to maximum white and the maximum negative difference to maximum black.

Only D_0 and D_4 are presented, because all other difference fields consist entirely of random brightness points. In the case of familiar stereo pairs, the difference field D_k contains points of near-zero value forming contour-lines having equal parallax shifts of k picture elements.

The next two pictures, in Fig. 29, show D_0 and D_4 for the case of the unmixed perturbation grid in Fig. 5. Through the perturbation grid, the uniformly gray background (or square, respectively) is clearly visible. Now, taking the mixed perturbation grid in Fig. 19, D_0 and D_4 should be very similar to the unmixed case. In the unmixed case the perturbation grid is always black (or always white) for D_0, which for the mixed case

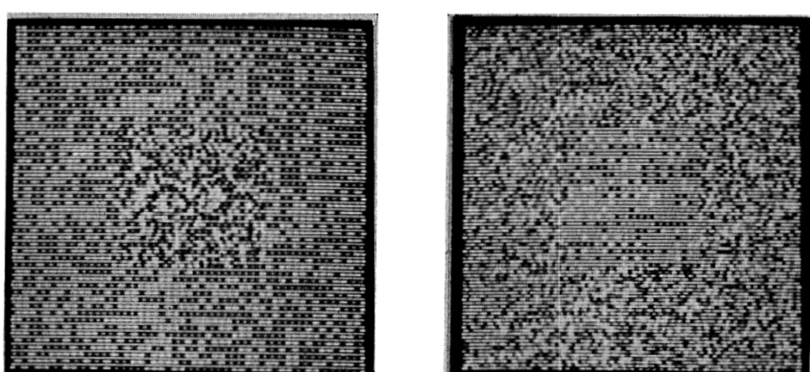

Fig. 30 — Difference fields D_0 and D_4 for the case of Fig. 19.

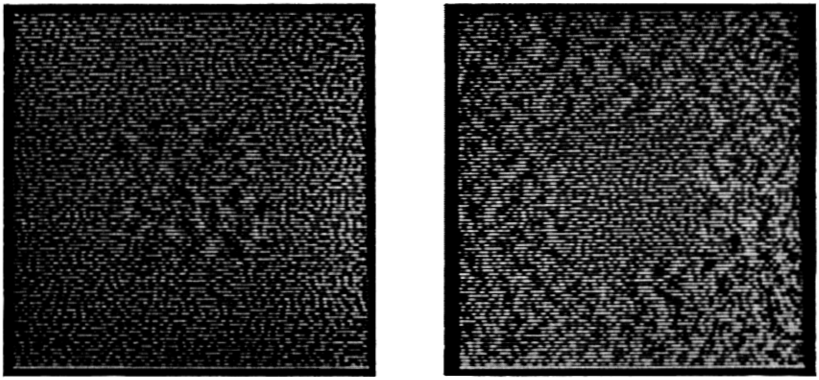

Fig. 31 — Difference fields D_0 and D_4 for the case of Fig. 13.

yields the same regular grid, but the grid points can take black and white at random. The left picture in Fig. 30 shows D_0 in this case, and it is now striking how well, in contrast to Fig. 29, the random central square is hidden by this type of perturbation. The right picture in Fig. 30 is D_4 for the mixed perturbation grid. Here, the grid points can take black and white values with 25 per cent probability each, and gray values with 50 per cent probability. Therefore, only 12.5 per cent of D_4 is effectively perturbed, but, because of the random appearance of this perturbation, it is more effective in hiding the central square than is 25 per cent perturbation of the unmixed grid. The uniform regions must be detected both in D_0 and D_4 to get depth.

In the next picture (Fig. 31), D_0 and D_4 are presented for the blurred picture of Fig. 13. The separation between the square and background is clearly visible, which confirms the fact that depth is also well perceived in this case.

By introducing gaussian noise perturbation in the stereo pairs (as in Fig. 11), D_0 and D_4 were determined. Subjective experiments were then conducted to determine the amount of noise that cancels depth, and this amount was compared with the noise needed to hide the square in the difference fields.

The results of this experiment, using ten subjects, indicated that the threshold of perceiving depth was 6 db signal to noise (with a very rapid decline in depth perception below this value), and that the *same* threshold value was obtained for the detection of the square in the difference fields.

As was emphasized before, the difference fields are probably very crude analogies for the binocular fields; nevertheless, it is worthwhile to mention the following fact: In the course of these investigations a great number of different perturbations were introduced in the stereo random fields. As a result of this process, the obtained stereo pairs could be rank-ordered according to stability and time required to perceive depth. This same ordering process was performed on the corresponding difference fields based on the separability of the central square and its surround. It turned out that the two established hierarchies were identical, except for borderline cases. Naturally, such subtleties cannot be explained by our simple analogy, especially if we consider the following: We performed monocular pattern recognition on the difference field in order to detect certain regions, while binocular pattern recognition was performed on the binocular field to get depth. There is no evidence that the laws of binocular pattern recognition are identical to ordinary (monocular) pattern recognition. (For instance, it is known that connectivity is an important monocular pattern-recognition cue that seems to be even more emphasized in binocular pattern recognition.)

Even the assumption of using a linear operation (subtraction) in the model is naturally an oversimplification. In the next experiment we demonstrate a nonlinear phenomenon of the binocular space. The perturbation grid in Fig. 32 is used. Here, every even sample in every even

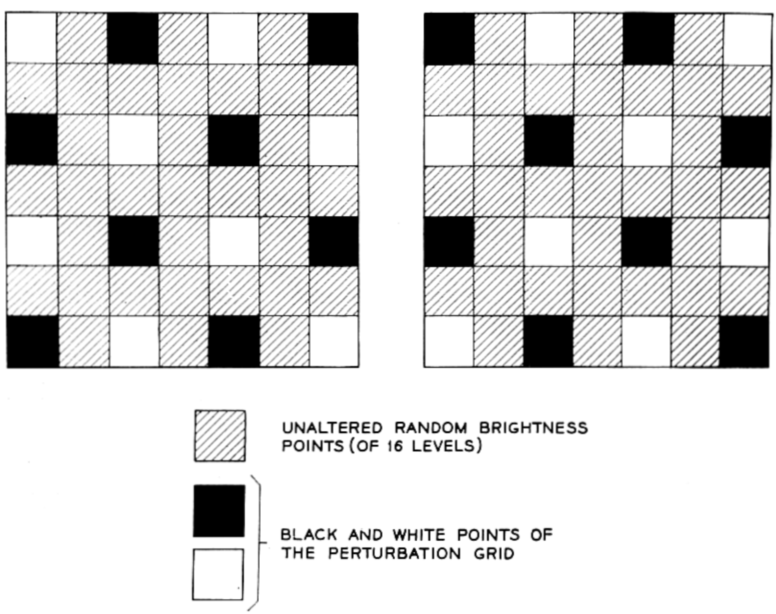

Fig. 32 — Illustration of the method by which the alternately mixed perturbation grid was generated.

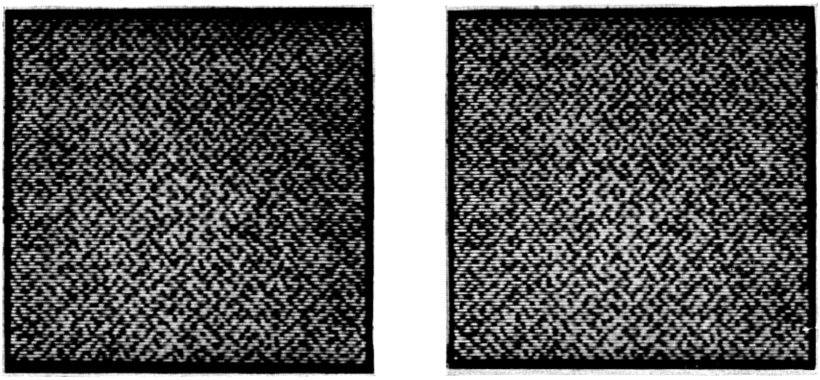

Fig. 33 — Stereo pair with alternately mixed perturbation grid.

line is alternatingly black and white, and its complemented value is copied in the other stereo picture. Fig. 33 demonstrates this case. Under strong illumination depth cannot be perceived. When the lights are dimmed (or eyes squinted), depth is easily obtained. Fig. 34 shows the difference fields; here, we find that detection of the center square is somewhat dependent on the illumination. However, this weak dependence is not consistent with the depth experiment.

X. MONOCULAR MACROPATTERNS ENHANCE DEPTH PERCEPTION

In posing our original problem we were interested in whether the perception of depth uses monocular pattern recognition, binocular pattern recognition or a combination of both. In the previous sections it was demonstrated that depth can be perceived without monocular patterns

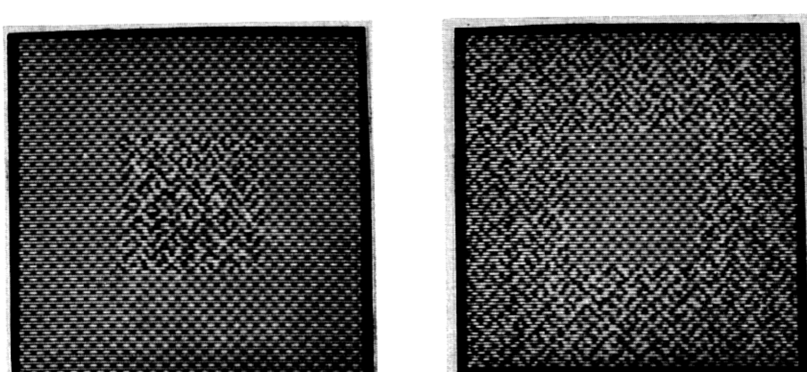

Fig. 34 — Difference fields D_0 and D_4 for the case of Fig. 33.

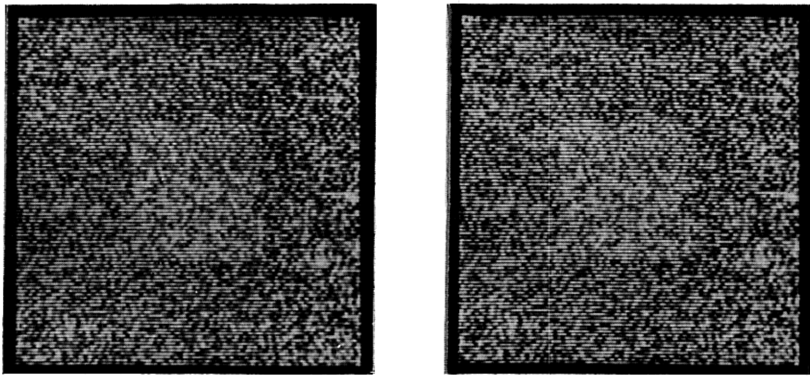

Fig. 35 — Stereo pair with brighter center square.

being present. In this section it will be demonstrated, nonetheless, that monocular macropattern recognition enhances depth perception. The same random stereo images are used, but the average value of the brightness points of the square is increased. Because of this, the random points in the square are brighter than the surround, and the square can be also seen monocularly. Fig. 35 demonstrates this case; it is apparent that the depth effect is obtained much faster than it is with missing monocular cues. According to this, we can suppose that depth perception is a combination of binocular and monocular pattern recognition, as was suggested in Fig. 3.

The actual processes of depth perception are, of course, much more complicated than the simplified diagram in Fig. 3. The different blocks are probably connected in many ways. Complicated feedback loops exist

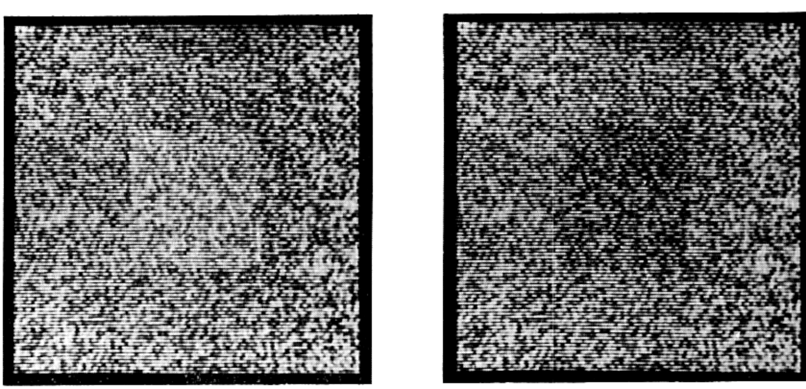

Fig. 36 — Stereo pair with whiter left and blacker right center square.

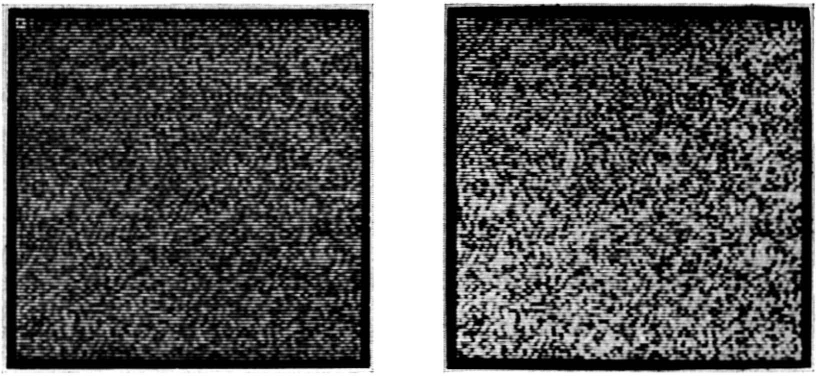

Fig. 37 — Stereo pair with left picture attenuated three times.

between the binocular field and the monocular fields, between the binocular pattern recognizer and the depth perceiver, etc. Fig. 36 demonstrates such a feedback between the binocular and monocular fields. Here, the random points in the left square have a mean value 20 per cent less than the surround and 20 per cent more than the surround in the right field. By fusing Fig. 36 we can see the square in depth with apparently the same brightness as the surround.

Fig. 37 shows another interesting case, where the left brightness values are attenuated by dividing them with a factor of three. In this experiment, $\Delta = 7$ picture elements and the center square is only 30×30. Depth is still easily perceived, according to expectation.[7]

Another even more complicated operation takes place in the monocular fields in connection with the binocular field. In Fig. 38 the left pic-

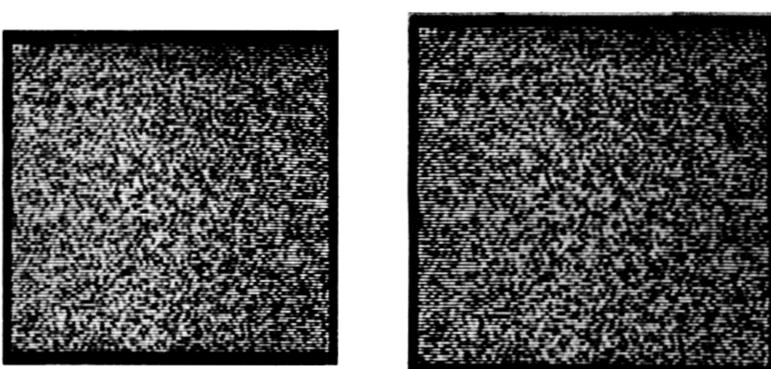

Fig. 38 — Stereo pair with right picture expanded by 10 per cent.

ture is contracted 10 per cent in both height and width. Even with this tremendous size discrepancy, fusion is possible and depth can be perceived. The same is true for rotations. More than ±6 degrees rotation from the base line can be tolerated and depth perceived.

The thorough investigation of these processes is the key to real understanding of depth perception. Some of the techniques developed here might be useful in such further exploration.

XI. SOME PROPERTIES OF THE DEPTH PERCEIVER

In Figs. 1, 2 and 3 the pattern recognizers were followed by a block called the "depth perceiver." This unit might have the function of coordinating several pattern-recognition tasks and assigning depth to various points. Even those points that have no parallax (seen by one eye only) will be located in depth. When there is no contextual reason to assign a particular depth to certain ambiguous point domains, there is a general tendency to see them in the farthest plane.

This tendency can be demonstrated by fusing Fig. 39. Here, the ambiguous random points lie in the place of the uniform black square seen behind the surround. Some investigations of ambiguous stereo effects (without parallax shift) were recently carried on with a similar result.[8]

The depth perceiver is particularly sensitive to any vertical shift (perpendicular to the base line). Parallax shifts with slight vertical components will not give rise to depth effects, probably because such shifts cannot occur in life. It seems reasonable to assume that the depth perceiver utilizes monocular depth cues too.

Naturally, all such divisions into different blocks are mere specula-

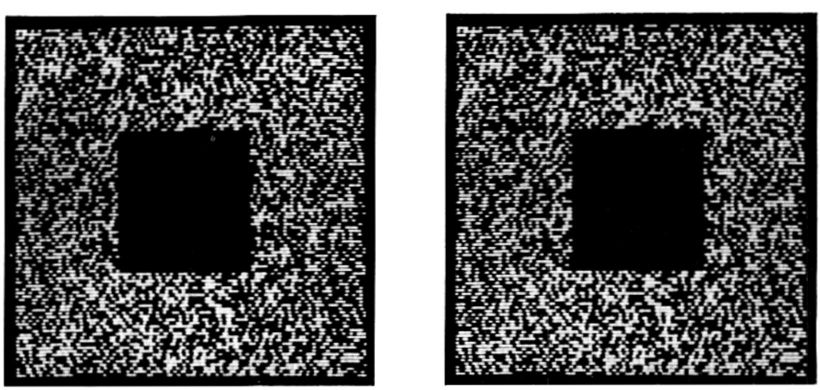

Fig. 39 — Stereo pair with uniformly black center square behind the random foreground.

tions until other psychological and physiological findings give adequate support.

XII. CONCLUSION

The peculiar depth effects that have been demonstrated strongly suggest that, under these conditions, depth perception is closely related to pattern recognition processes on the binocular field. Someone could raise the question: What is the merit of showing that binocular and not monocular pattern recognition is required in depth perception if the processes of pattern recognition are still unknown?

To answer this, we must realize that pattern-recognition processes are complex and highly nonlinear in nature. Because of this, it is very important which operations are performed on the input patterns before recognition. (For instance, upon performing the pattern-recognition task on the difference fields of Fig. 29 and Fig. 30, the qualitative difference of perceiving depth in the two cases is instantly apparent, which could not be simply explained if the recognition had been performed on the monocular patterns of Fig. 5 and Fig. 19.)

Thus, the discovery of certain transformations of the input patterns that facilitate the recognition task provides better understanding of the laws of pattern recognition.

These experiments indicated also that, without monocular cues or *Gestalt*, depth can be still perceived. In order to be seen in depth, the patterns need to possess much simpler properties (e.g., one-dimensional connectivity, adequate number of connected points, etc.) than we originally expected. These properties might be simple enough to be simulated by present computer technology. Thus, the findings of this study might give a new impetus to the development of devices that will determine depth automatically.

The technique of stereo random fields also has several advantages in a great variety of possible applications. In binocular fusion studies, the problem of binocular rivalry sometimes makes investigation cumbersome. These stimuli have a self-checking feature against binocular rivalry; namely, as long as depth is seen, no rivalry can be present.

The long time constants needed to perceive depth in certain presentations indicate that depth perception depends very much on the input material. From the order of a few milliseconds (required for simple stereo pictures), we can easily increase the perception time to the order of minutes. This slowing down of a process can be very advantageous in investigations of learning, pattern recognition, etc.

The stability of the random stereo fields is also very useful. Because

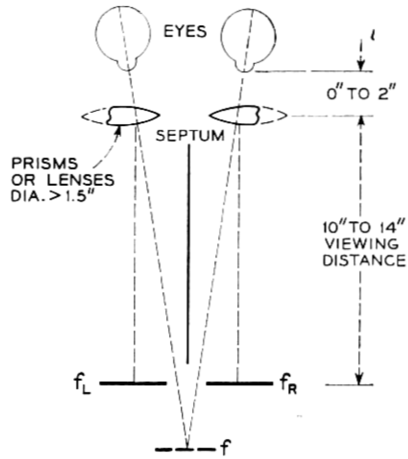

Fig. 40 — Illustration showing how presented stereo pictures should be viewed.

nearly all points carry depth information, the stereo image is very stable and points with greater parallax shifts than in the ordinary case can be fused.

Such stimuli could also possibly be used in apparent motion studies.

This technique was found to be a useful tool in color studies to examine the role of color in depth perception.

But perhaps the most useful property of this method is the elimination of context and higher organization from the input stimulus, which makes it possible to isolate and study less formidable problems.

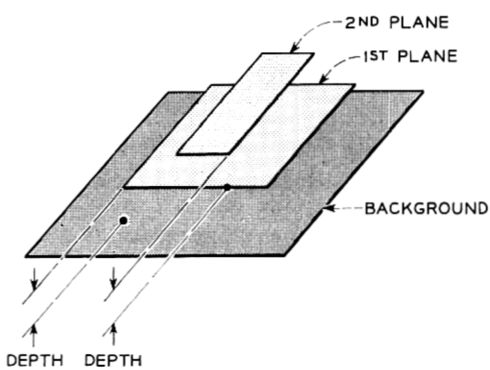

Fig. 41 — The subjective illusion seen when Fig. 8 is viewed stereoscopically.

XIII. ACKNOWLEDGMENT

I am indebted to E. E. David, Jr. for his valuable comments on this paper and to R. A. Payne for his skillful assistance in taking and processing the many photographs.

APPENDIX

The presented stereo pictures can be fused if they are viewed through a pair of lenses used as prisms, as shown in Fig. 40. The focal length of the lenses should be 10 to 18 inches and their diameter around $1\frac{1}{2}$ inches or more, as is the case with the ones accompanying this paper. Sometimes it takes several minutes to get the depth effect.

If fusion of the left and right images cannot be obtained easily, a stiff paper or cardboard septum (10 to 14 inches long) placed between the two stereo pictures and perpendicular to the page will probably eliminate the difficulty (see Fig. 40). Viewers who ordinarily wear glasses should not remove them when using the lenses.

For example, the subjective illusion that is seen when Fig. 8 is viewed stereoscopically is illustrated in Fig. 41.

Paste envelope here,
flap down and
to the right

REFERENCES

1. Gibson, J. J., Perception of Distance and Space in the Open Air, Army Air Force Program, Report No. 7, 1946, p. 181; reprinted in Beardslee, D. C. and Wertheimer, M., eds., *Readings in Perception*, D. Van Nostrand Co., New York, 1958.
2. Langlands, H. M. S., Experiments in Binocular Vision, Trans. Opt. Soc. London, **28,** 1926, p. 45.
3. Wheatstone, C., On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision, Roy. Soc. London Phil. Trans, 1838, p. 371.
4. David, E. E., Jr., Mathews, M. V. and McDonald, H. S., Experiments with Speech Using Digital Computer Simulation, I.R.E. Wescon Conv. Rec. — Audio, August 1958, p. 3.
5. Graham, R. E. and Kelly, J. L., Jr., A Computer Simulation Chain for Research on Picture Coding, I.R.E. Wescon Conv. Rec.—Computer Applications, August 1958, p. 41.
6. Julesz, B., A Method of Coding Television Signals Based on Edge Detection, B.S.T.J., **38,** 1959, p. 1001.
7. Ogle, K. N. and Groch, J., Stereopsis and Unequal Luminosities of the Images in Two Eyes, A.M.A. Arch. Ophthal., **56,** 1956, p. 878.
8. Wegner, K., Stereoskopische Untersuchungen für Tiefenlokalisation sogenannter funktionsloser Bestandteile des Binocularen Gesichtsfeldes, Thesis, Georg-August Univ., Göttingen, Germany, 1959.