International Journal of Computer and Information Sciences, Vol. 3, No. 2, 1974

On an Information Processing Model Applied to Optical Illusions and Computer Simulations

C. A. Mateescu[1] and I. O. Stamatescu[2,3]

Received April 1973; revised December 1973

A model for image processing proposed earlier is discussed and applied to some optical illusions. A simulation on a computer is provided in an attempt to reach a quantitative level of understanding in the study of visual phenomena.

1. INTRODUCTION

Attempts have been made over the past decade to develop information processing models for some visual phenomena related to the appearance of optical illusions.[1–3] At one time most of the features of the human perception of images were thought of within the framework of "synthetic" theories of the kind of Gestalt theory.[4] A "macrostructure" for perception was provided, introducing the concepts of continuity, total shape, good form, etc. Subsequent research in this field and connections with information theory showed, however, that such synthetic concepts could and must have a structure themselves. This appears in a natural way in a picture that represents "visual activity" (we shall use this phrase to indicate the whole

[1] București, str. B. Delavrancea 27.

[2] Institut für Theoretische Physik der Universität Heidelberg, Heidelberg, Germany.

[3] Preliminary work on this subject was done while this author was at the Institute of Physics of CSEN, București, Romania.

[4] See Ref. 4 for a review of theories of perception, as well as new developments and original research.

process by which images are perceived and interpreted) as a highly intercorrelated process with a primary statistical basis.[1,5] This picture is assumed more or less consistently in most of the studies[1-3,5-7] dealing with optical illusions.

Optical illusions provide well-defined examples for examining what is thought to be characteristic of the perceiving of images in general. Some approaches sum the correlation effects and describe them formally by potential functions.[6] By comparing the integrated mutual interaction of forms with experimental observations one could characterize these potentials. Though elegant, this line of attack has a shortcoming arising from the difficulty of reassigning a physiopsychological meaning to the features of these potential functions if they are not of such a simple form as, for instance, those corresponding to the "centration effects" described by Piaget.[5] Other approaches try to uncover elementary mechanisms which produce deformations of the contours.[2,3,7] These models, remaining at a qualitative level, have the problem of achieving definite verification, since counterexamples can be ruled out very often by speculating about nuances in the results. In any case we feel that the problem cannot be consistently studied without first elaborating a clear scheme for the processing of visual information which can allow us to go beyond a qualitative level.

In this paper we approach the problem of a quantitative analysis of visual activity. The great amount of information processed and the great number of correlations established in very short times suggest the use of statistical concepts. The frame of the analysis is given by a model proposed in Ref. 8. We shall discuss the specific features of such a program and then, concentrating on optical illusions, we provide an example of computer simulation of the process by which the distortion appears in the Herring illusion.

The study we perform here is to be placed in the general framework of functional theories.[1,5] It does not attempt a mechanical simulation of the functioning of the physiological structures. The experimental data on the physiology of the retina lead to a more or less comprehensive picture of its functions.[9] However, with regard to the higher pathways and the central system, the present data give a measure of the complexity of their organization[10] but do not allow an attempt to decode it. A direct simulation cannot therefore be undertaken at present if one is interested in the phenomena of real visual activity, which, as is usually the case, cannot be related to a single physiological background (the retina).[5]

Besides this constraint, however, we also think that the use of statistical

[5] An example of an effect explanable on the basis of retinal functions is that of the Mach bands.[9]

methods in the framework of information processing models[1,5,6] is consistent with one of the characteristics of the objective process. On the basis of the physiological data one could consider that nature does not work on minimal schemes but by an accumulation of facts leading to enormous overcomplexification. The essential differences between the information received by the eye and the information at the conscious level (in amount, quality, and manner of treatment) make it tempting to assume that statistical functions are characteristic of the visual activity itself. This supposed compatibility of the instrument and the object could lead to the degree of simplicity ensuring understanding.

The verifiability of this kind of model relies not on its analogy to the functioning of a neural network at each step, but on its capacity to reproduce complex phenomena occurring in vivo. The insight one can hope to gain thereby concerns, first, psychological understanding and then, indirectly, physiological understanding, by suggesting the hierarchy of the structuring mechanisms and their essential features.

2. GENERAL HYPOTHESES ABOUT THE PROCESSING OF IMAGES

In Ref. 8 a model was proposed which considered three levels for the construction and interpretation of an image (Fig. 1). At the first level ("the level of visual stimuli") data are collected from the image-object, the eye focusing different parts of it. A first filtering of the information appears, following the conspicuousness of the points, but otherwise to a great extent at random if we do not take into account the influence from the other levels (see below). Retinal inhibition[11] could account for some distortions[3,11] and retinal stabilization[2] could also determine specific effects at this level.

At a second level (called in Ref. 8 the "perceptual level") equilibrium and continuity criteria induce a recognization of patterns. Then the whole visual field is organized in a structure in which usually a few of these patterns play a determining role.[8] Centration effects[5] and in general "field effects"[5,6] are bound to manifest themselves at this level.[6] (By extension, in Ref. 8 all these effects were called "centration-type effects.")

These levels are already closed by two feedbacks. One affects the filtering of data at the first level, which will disregard what does not fit into recognized

[6] We understand here by field effects any effects due to the global structure as taken in contradistinction to the local structure. They have, however, various origins and usually show extensive features with respect to their composition. The recognization of patterns usually induces field effects.

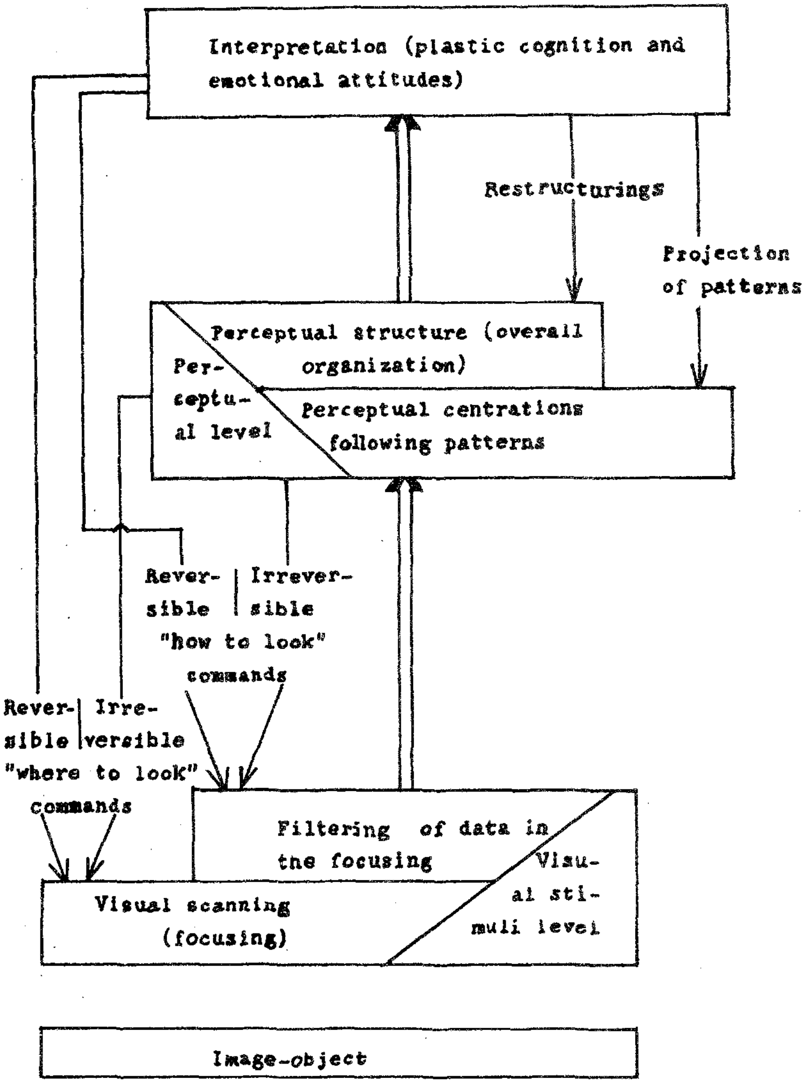

Fig. 1. A model for image processing.

patterns as long as these themselves form a consistent structure.[1] The second affects the scanning of the image in the sense that some particular regions get a greater weight in the "field of attention." These feedbacks, which we shall call of type I, induce essentially irreversible effects[5] leading to a stationary process, because they are controlled by the peculiarities of the figure itself as it appears to a given subject.

At the final level an understanding of the image is attempted (together with which a "response" is prepared). Any interpretation (including comparisons, etc.) is subjected to the inhomogeneities of the perceptual structure provided by the second level. It seems to us merely a matter of speculation to decide upon whether the perceived image is deformed following the perceptual structure, or whether the metric to which we subject the visual field is modified accordingly.[1] Independently, however, of whether we assume a homogeneous interpretation of a structured perceived image, or a structured topology of the interpretation of the real figure, it is clear that the structure itself is well defined and therefore we considered a definite level at which it is established (the second level).

The feedbacks from the third level affect both of the preceding levels; for the use of perceptual "activity" patterns are projected at the second level, and are no longer merely "well-balanced" forms but involve a certain understanding and experience[5] (perspective, human shape, etc). Moreover, they command possible restructurings of the whole "perceived image" (perceptual organization), according to the significance found in the forms (type II feedback). On the other hand, they influence directly the scanning and the filtering at the first level, merely in a conscious attempt to lower the degree of irreversibility introduced in the first part of the process (type III feedback).

Rescaling, renormalization, and other "corrective" actions are induced from this third level, either directly (type III feedback) or indirectly (type II feedback). Through restructuring, the indirect action can provide perceptual shifts (as, e.g., in the perspective-type illusions). The direct action does not usually produce such shifts, aside from determining the alternation of different structures in the case of more complex figures (e.g., the Necker cube, impossible figures[1]). However, if instabilities in the perceptual structure allow a "pseudoreversibility" to be reached, the direct action from the third level can correct possible tendencies to distortions. This pseudoreversibility implies destructive interferences of the field effects and of the type I feedback and allows the type III feedback to induce compensations of the shifts. In the case of illusions the type III feedback fails to correct the image. The structure is very stable and the elementary deformations are enhanced by the strong irreversibility.[7]

Though it constitutes only a simplified description, this model evidently incorporates some of the main features of the visual activity, providing a consistent scheme for its decoding.

We want to emphasize that the scheme we have described is an opera-

[7] In the case of impossible figures the interference does not cover the whole image, although the patterns are not compatible. Due to the strength of the perspective effects (coming through type II feedback) the structure encounters a periodic restabilization, producing a spatial curvature or torsion when the image is small enough to be focused globally.

tional scheme and its elements are not to be directly related to the physiological background. Nor could an attempt to find their physiological realizations be fruitful at present. Since no function of the operational scheme is supported by a single aspect of the physiological background, it does not help to try to decode parts of the scheme in terms of the known functions of the neural networks.

On the other hand, existing physiological information is taken note of by our model. In some of the elementary functions we recognize, for instance, the background inhibition at the retinal level[9] or the identification of line elements at the central level (suggested by the slit form of the receptive field of simple field cells[10]). Moreover, the hierarchy of the interconnections and the loss in pointlike specificity accompanied by the increase in (what we could call) figural specificity in moving to higher levels[10] are in agreement with basic assumptions in the construction of our model.

The picture described above involves a continuous activation of the ascending and feedback paths. The basic randomness of the process is modulated by probability distributions reflecting the recognization of patterns and the assignation of meanings, which increase the speed of the analysis. The feedback induces an optimization of the scanning, which proceeds further on a "trial and error" basis, as suggested in Ref. 1: After a reduced number of elements signalizes a pattern, the subsequent scanning follows with great probability the lines predicted by the pattern.

The scanning itself has a quite complicated nature.[2] It probably can be understood operationally as consisting of two stages, the realization of the retinal image, and its further analysis, as suggested by the after image effects.[11] These two stages are characterized by different processing speeds, degrees of seriality, and decay parameters. Moreover, they are intimately interconnected and by no means clearly separated in time in the case of normal visual activity. That is why a seriality of the scanning cannot be well defined. In focusing on a point, information from far away points is also present and only statistical considerations can work in a stratified scheme which accounts for both local and global effects.

The decoding of the scanning function involves phenomena occurring at the retinal level, on the higher pathways, and in the central system. It includes, e.g., variation of contrasts[9] and continuity criteria.[4,10] These are, however, only primary phenomena, since the brain's hardware also contains higher configurations using simple field units as building blocks.[10] We feel that by using a statistical framework to account for the scanning we can, at least roughly, simulate the effect of the higher degree of inner connectedness peculiar to it. The same remarks are in order for all the other elements of our scheme.

Stable structures are reached when the cybernetic scheme is equilibrated, i.e., when the process becomes stationary. The average frequency of processing at each point becomes constant in time. The structure of the probabilistic distributions as functions of the points and of the directions (the latter giving specific "serialities") acquires an inhomogeneity and an anisotropy over the field of the image which is preserved from now on. Usually the stationarity is reached very rapidly and it is followed by the stabilization of the perceptual structure. With the increasing complexity of the image, different stable solutions may exist, which could alternate or share parts of the image (e.g., the Necker cube, impossible figures, "hidden" figures).

Of course, the situation is far more involved for really complicated figures—for instance, paintings. As was pointed out in Ref. 8, the perceptual structure for a painting is so rich that the fluctuation in time of the inhomogeneities and anisotropies becomes itself relevant. One could imagine a very great number of "metastable" structures[(12)] whose alternation in time and space generates rhythms and coordinations on which the real understanding of the painting is grounded. The figurative-literary element, if it exists, is merely auxiliary—it is not the fundamental element. The profound impression, the true meaning comes through complex mechanisms involving a direct action of these rhythms, coordinations, tensions, etc.[(8,12,13)] This type of information is specific to the visual arts. It can be associated or subordinated to a "literary symbol" as in certain figurative paintings, or it can constitute by itself the "message."

Taking into account a decay in time of the information, a region will be the more persistent—and thus pregnant with significance—the greater is the average weight of its point in the attention field. Consequently, the strength of the effect generated in this region depends on these weights, being greater when the number and the density of the strongly weighted points is greater. On the other hand, effects due to multiple correlations have a lower rate of accumulation, due to the fact that their stabilization requires repeated analysis over a wider region. If an effect appears as the result of correlations between the information from many elements, then the average loss of information at each element increases with the number of elements, because of the corresponding increase in the average time separation between two processings over the same element. However, if the effect is supported by a pattern (already existing in the perceptual structure, or induced by the cumulative effect itself), the corresponding region will be preferentially weighted, alterations facilitated, and the effect will be increased. Once the threshold for the recognization of the pattern is reached these effects are accounted for by the perceptual structure. We remark that the accumulation of information cannot grow indefinitely. It reaches a certain threshold which constitutes a second degree of information, which is further processed.

3. SIMULATION ON COMPUTER FOR A MODEL DESCRIBING SOME OPTICAL ILLUSIONS

Some time ago Simon[1] proposed a model for impossible figures which involved a serial processing of visual information. We saw in Ref. 8 that the seriality which can be found at the different levels of visual activity is by no means as simple as, for instance, in reading.[8] We tried to come closer to what has been assumed for the actual process by considering a statistical processing of small but finite regions of the field, following various distribution functions, with precocious recognization of simple elements as lines, etc., where the distributions can account for global features. This implies a rather rough description of the complex process mentioned above; we feel, however, that in the context of the presently available information this approximation is not too bad.

We shall fix our attention on the example of the Herring illusion (Fig. 2a). We first remark that the illusion disappears in Fig. 2(b), while persisting in Figs. 2(c. d). This indicates that the essential role is played by the intersections. Moreover, it has been observed that the effect is cumulative and that the outer intersection angles have to be smaller than a certain value (around 30° [8]) in order that the illusion appear.[3] There are of course other possible explanations for this figure.[2,3,6,7,11] We consider that in this and other cases the elementary effect described in point 1 below is dominant.[9] With this specification the mechanism we assume for optical illusions is the following.[10]

1. The elementary shift appears as the result of an ambiguity in the scanning when reaching an intersection. At the place where two lines come together the points near the vertex are somewhat mixed up and there is an ambiguity in following one or the other line. A first result of this is a slight shift of the vertex toward the interior of the smaller angle (where the effect is greater; no displacement appears for right angles, the effects being small and reciprocally compensated). The degree of ambiguity is influenced by asymmetries in the surrounding structure and the strength of the effect, depending on the statistical weight of the region, will be decreased, for instance, by shortening the lines. We consider that this effect is partially responsible for the Müller–Lyer illusion. If previously a direction was being followed, a second effect appears from a tendency to equilibrate the angle with respect to that direction. Any asymmetry in the local or global structure ensures the non-

[8] This can be estimated looking at the figure from one side.

[9] Most of the explanations, however, can be made with some agreement to the data.

[10] There is some indication that the illusion is not a retinal effect, suggested by an informal experiment where one presents to one eye the obliques and to the other the horizontals. The authors thank Drs. I. Bender and D. Gromes of the Institut für Theoretische Physik Heidelberg for this remark.

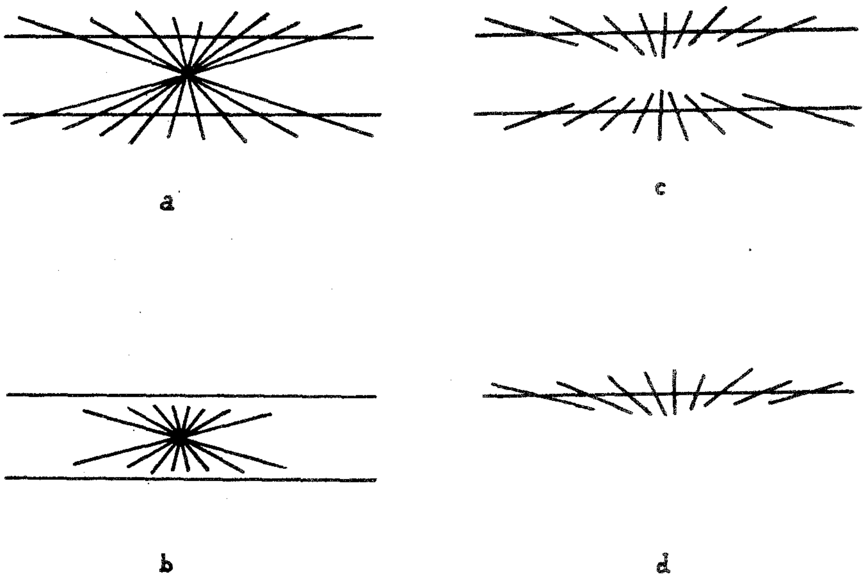

Fig. 2. (a) The Herring illusion. (b) The inner part is not essential. (c) The essential part. (d) The simplest figure which shows the distorsion.

compensation of the shifts and induces a displacement which shows a maximum for medium angles (with increasing angle the shift of the line increases, but that of the vertex and the ambiguity allowing them decreases). The two effects are indicated in Fig. 3.

2. The elementary shift itself usually gives only a small perturbation, since any apparent deviation is recognized and corrected by the eye. If, however, a sequence of such shifts exists, their cumulative effect creates some kind of a "false" pattern which, by way of feedback, induces an irreversible shift. The appearance of this false pattern strengthens the asymmetries which support the shifts, on the one hand, and on the other hand it restructures the field on new equilibrium lines. Certain features of the global image can also produce such "false" patterns—for instance, the rescaling in perspective-type illusions. The result is that if the constructive interference of the shifts

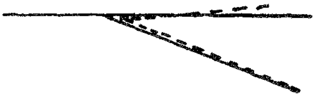

Fig. 3. The elementary effect as assumed in the analysis (emphasized).

reaches a certain threshold, they will be enhanced by the normal mechanism of visual activity. A stable deformed structure appears.

3. If superficial figures are well defined, the statistical weights are normalized not only with respect to lengths but also with respect to areas. Since we will not be returning to this point later, we remark that in the case of the Delboeuf illusion the areas involved are those of the inner circle and the ring, which are nearly equal for a ratio of the diameters of 3/2. In the corresponding afterimage experiment, however, the areas of the two circles intervene. The elementary shift itself is probably due to a centration effect but the mechanism responsible for the stabilization of the illusion is more complicated, including, for instance, comparisons.

In the case of the Herring illusion the asymmetries which help the constructive interference are of two kinds: (i) because of the special status of the horizontals, most of the effects are refered to them (this is also valid for the Zöllner figure), and (ii) the (actual or potential) concurrence of the obliques brings an up–down asymmetry (Fig. 2d).

The last point of our program is to put a model for the illusion into a numerical form in order to see quantitatively how the different mechanisms act. To save computer time, we considered scanning on the horizontal only, with deviation on the obliques corresponding to the mentioned ambiguity, together with which small shifts are given to the segments of the horizontal. Points are picked up statistically and small regions around them are investigated at one time. Deformations of the obliques are minor anyway, since no cumulative effect appears. Some feedback is explicitly included to induce directionality if no order was observed for a long time (and conversely; only type III feedbacks), and also to reevaluate the degree of ambiguity at the intersections and other probability distributions. The type I and II feedbacks are accounted for also in the initial distributions. The rate of the decay of the false information is reduced when the cumulative effect reaches a given threshold.

The computation performs about 1000 processings per point, the process being implemented by a statistical average over the shifts leading to the distorted line presented on Fig. 4. The logic scheme and details about the program are given in the appendix.

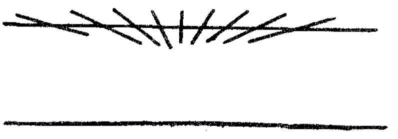

Fig. 4. The result of the calculation (Fig. 2d is reproduced to allow comparison).

4. DISCUSSION

At the level of this work the computation we present proves that the model works but does not establish once and for all the truth of our assumptions. To come nearer to this point, we would have to consider a simultaneous fit of many similar problems and using the experimental information find as output the parameters entering the weights, the shifts, and the decay functions. Nevertheless, the stability of the structure obtained makes the computation represent more than a mere illustration and numerical realization of the model. In any case it indicates a method for simulation by computer of visual phenomena.

Some of the implications of our assumptions can, however, be directly tested. The deformed structure is the result of the establishing of many correlations and before a certain threshold is reached the false information decays very rapidly in time. We therefore conjecture that if the figure is seen intermittently (through a rotating disk with a cut, for instance), such that the "open" intervals are small enough not to allow many processings to be performed (and thus the enhancement mechanism to start) and the "closed" intervals long enough to ensure a practically total decay of the memorized shifts, the illusion will tend to disappear. The corresponding durations are probably of the order of milliseconds and tenths of a second, respectively. By varying these intervals, one could analyze the features of the process. Connections can be made with research in visual information media.

We would like to close this discussion with the remark that research during the last few years has given us the hope that the complex process of vision could be analyzed and its features displayed. This would provide, then, a basis for the understanding of the more complicated problem of visual communication, which is fundamental for the study of the art. Such an approach to the problem of art was tried in Ref. 8, where the semiotic system of the visual arts was investigated. Earlier studies already signaled some of these features[12,13] sometimes proceeding from a more literary and speculative point of view.[14] General problems in the framework of a quantitative approach to aesthetics are treated in Ref. 15.

APPENDIX

The program was written in FORTRAN IV G and run on a high-speed IBM 360 (DOS) computer. The logic scheme is presented in Fig. 5. The length of the horizontal is 100 points. The information from the shifted elements is cumulative and decaying in time (here counted after the number of points analyzed). Memory of the analyzed points is carried along to permit feedback. They induce from time to time a certain degree of order, a

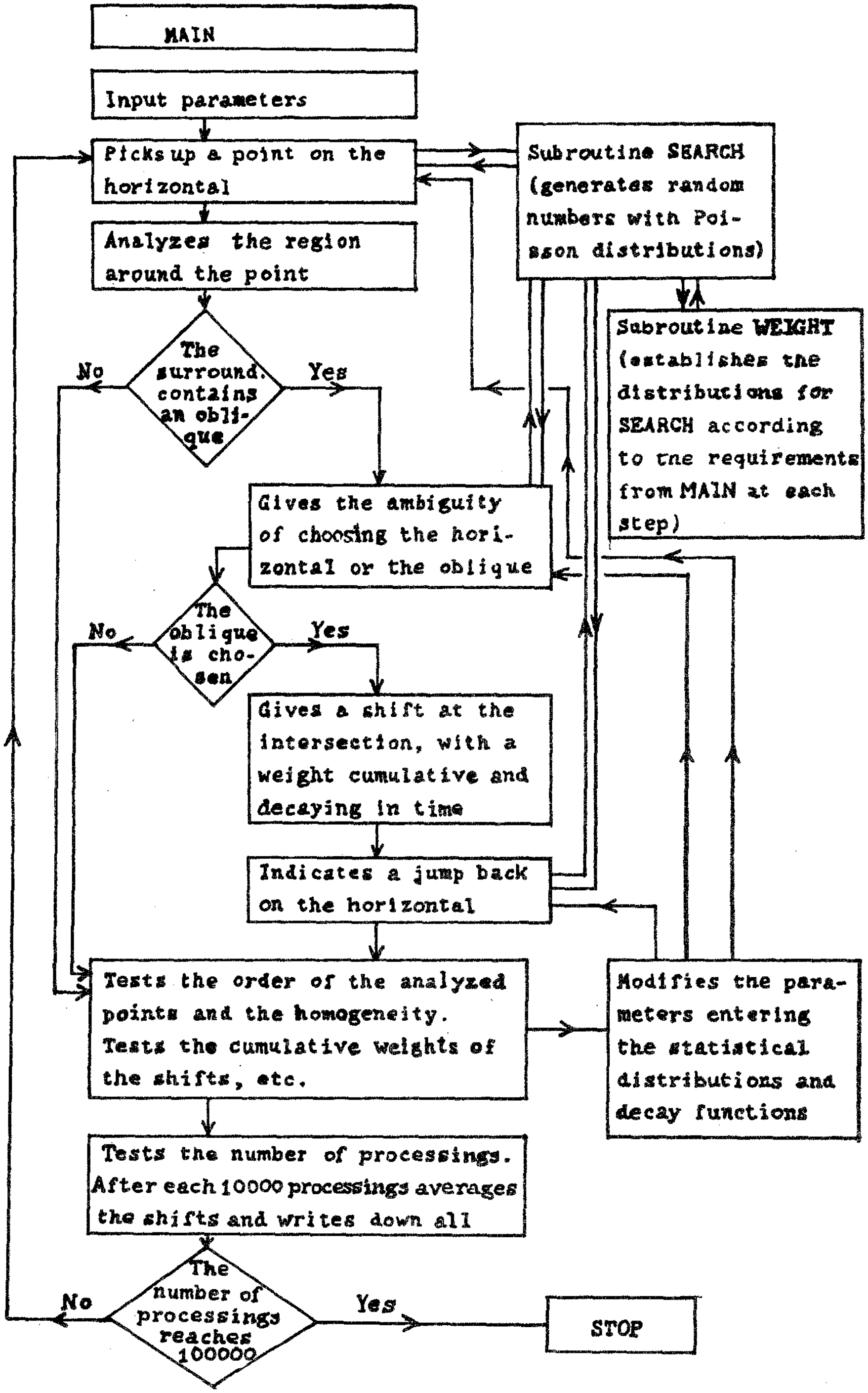

Fig. 5. The logic scheme of the FORTRAN Program.

reduction of the asymmetries, or a variation of the degree of ambiguity, to simulate the action of the type III and of some of the type I feedbacks. They work by modifying the parameters of the probability distributions at various steps (the distributions are throughout assumed to be of the Poisson type, though of course there is no clear indication for this). The feedbacks also control the decay in time of the information from the shifts, slowing the decay when the accumulation reaches a given threshold (type II + I feedbacks). The recognition of the false pattern induces the new equilibrium lines corresponding to the distorted figure. At this point an average over the shifts is performed, and the result is written down. The computation showed that the deformed image is stabilized before half of the running time elapsed. Since no fit was tried, the initial parameters were chosen by convenience, according to our assumptions. It is to be remarked that a good separation of the different effects can always be made. The qualitative structure is essentially dependent on the general assumptions. The strength of the deformation depends on the elementary shifts and thus on the degrees of ambiguity and asymmetry. The stabilization time depends on the decay rates and pattern recognization thresholds. The real computer time was of the order of 1 min for a total of 100,000 processings.

ACKNOWLEDGMENTS

The authors are indebted to Prof. S. Marcus of the Institut de Matematica Bucureşti for encouragement and valuable discussions. One of the authors (I.O.S.) want to thank Prof. G. Dosch of the Institut für Theoretische Physik Heidelberg for his interest and remarks.

REFERENCES

1. D. P. Andrews, "Error-correcting perceptual mechanisms," *Quart. J. Exp. Psychol.* **16**:104 (1964); H. A. Simon, "An information processing explanation of some perceptual phenomena," *Brit. J. Psychol.* **58**:1 (1967).
2. C. R. Evans, "Further studies on pattern perception and a stabilized retinal image," *Brit. J. Psychol.* **58**:315 (1967).
3. J. O. Robinson, "Retinal inhibition in visual distortion," *Brit. J. Psychol.* **59**:29 (1968).
4. E. Fischbein, *Conceptele figurale* (Ed. Acad., Bucureşti, 1963).
5. J. Piaget, *La psychologie de l'intelligence* (A. Collin, Paris, 1947).
6. R. S. Eriksson, "A field theory of visual illusions," *Brit. J. Psychol.* **61**:451 (1970).
7. G. H. Fisher, "Toward a new explanation for the geometrical illusions," *Brit. J. Psychol.* **60**:179 (1969).
8. C. A. Mateescu, "A quasilinguistic treatment of the semiotic system of the visual arts," to appear in *Poetics* (The Hague).
9. F. Ratliff, "*Mach Bands: Quantitative Studies on Neural Networks in the Retina*" (Holden-Day, Amsterdam, 1965); F. S. Werblin, "The control of sensitivity in the retina," *Sci. Amer.* **1973** (January):71.

10. D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.* **160**:106 (1962); D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture in two nonstriate visual areas of the cat," *J. Neurophysiol.* **28**:229 (1965); M. Kalia and D. Whitteridge, "The visual areas in the splenial sulcus of the cat," *J. Physiol.* **232**:275 (1973).
11. L. Ganz, "Lateral inhibition and the location of visual contours," *Vision Res.* **4**:465 (1965); J. L. Brown, "Afterimages," in C. H. Graham, ed., *Vision and Visual Perception* (Wiley, New York, 1965); H. L. Wagner, "The illusions and Ganz's theory of contour displacements," *Brit. J. Psychol.* **59**:361 (1968).
12. C. A. Mateescu, "Sur un espace selfconsistent en peinture," *Astra* **1968** (July).
13. N. Mouloud, *L'espace de la peinture* (Press, Univ., Paris, 1964), M. Schapiro, "On some problems in the semiotics of visual art: field and vehicle in image-sign," *Semiotica* §:3 (1969); M. Raphael, *The Demands of Art* (Princeton Univ. Press, 1968); I. Uspenschii *Poetica compozitii* (Istkustva, Moskva, 1970).
14. J. L. Schefer, *Scénographie d'un tableau* (Du Seuil, Paris, 1969).
15. S. Marcus, *Poetica matematica* (Ed. Acad., București, 1969).