

AD No. 402020

ACTM FILE COPY

JPRS: 18,126

14 March 1963

OTS: 63-21322

FOR REFERENCE ONLY AT EACH OF THE
DDC OFFICES. THIS REPORT CANNOT
BE SATISFACTORILY REPRODUCED; DDC
DOES NOT FURNISH COPIES.

DIGITAL COMPUTER RECOGNITION OF VISUAL IMAGES

by M. A. Ayzerman

- USSR -

APR 18 1963
JISIA

U. S. DEPARTMENT OF COMMERCE
OFFICE OF TECHNICAL SERVICES
JOINT PUBLICATIONS RESEARCH SERVICE
Building T-30
Ohio Dr. and Independence Ave., S.W.
Washington 25, D. C.

Price: \$.50

FOREWORD

This publication was prepared under contract for the Joint Publications Research Service as a translation or foreign-language research service to the various federal government departments.

The contents of this material in no way represent the policies, views or attitudes of the U. S. Government or of the parties to any distribution arrangement.

PROCUREMENT OF JPRS REPORTS

All JPRS reports may be ordered from the Office of Technical Services. Reports published prior to 1 February 1963 can be provided, for the most part, only in photocopy (xerox). Those published after 1 February 1963 will be provided in printed form.

Details on special subscription arrangements for JPRS social science reports will be provided upon request.

No cumulative subject index or catalog of all JPRS reports has been compiled.

All JPRS reports are listed in the Monthly Catalog of U. S. Government Publications, available on subscription at \$4.50 per year (\$6.00 foreign), including an annual index, from the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C.

All JPRS scientific and technical reports are cataloged and subject-indexed in Technical Translations, published semimonthly by the Office of Technical Services, and also available on subscription (\$12.00 per year domestic, \$16.00 foreign) from the Superintendent of Documents. Semiannual indexes to Technical Translations are available at additional cost.

DIGITAL COMPUTER RECOGNITION OF VISUAL IMAGES

- USSR -

[Following is the translation of an article entitled "They Teach a Machine to Recognize Visual Images", by M.A. Ayzerman in the Russian-language publication Nauka i Zhizn' (Science and Life), No 12, December 1962, pp 34-39.]

A session of the Division of Biological Sciences of the USSR Academy of Sciences devoted to problems of biocybernetics was held in Moscow in April of 1962. During the meetings major attention was given to the problem of machine simulation of various psychic functions of humans and animals. A particularly interesting report on this subject was presented by Professor M.A. Ayzerman, Doctor of Engineering Science, of the Institute of Automation and Remote Control. The following is an adaptation of this report prepared especially for this magazine by the author.

STATEMENT OF THE PROBLEM

The investigations, conducted by Engineer E.M. Braverman with the aid of Engineers O.A. Bashkirov and I.B. Muchnik, which I shall discuss are the first stage in a major joint study directed by the laboratory which I head and the Laboratory of Biocybernetics headed by Professor S.N. Braynes (of the A.V. Vishnevskyy Institute of Surgery of the USSR Academy of Medical Sciences).

The objective of the entire study is to attempt to approach an understanding of the mechanism of learning to recognize visual images and to develop the simplest possible methods for the reproduction of this process on general-purpose digital computers.

When we speak of learning to recognize visual images we have in mind the following two different learning processes which are well known to all of us from our

individual experience.

When the school teacher wants to teach a child to differentiate a triangle, for example, from other geometric figures he suggests that the student count the number of angles in the figure, that is he gives the student a plan of action -- a recognition algorithm. The student remembers this plan and with its help recognizes figures later shown to him. This understanding of the term "learning" is not of interest to us since if the algorithm of a recognition process is known then there is no difficulty in simulating it.

But there is another learning process associated with an ability of the human mind which we do not yet understand. Consider the following example. If we take many different shapes of the letters a and b and show them to a student who does not know the letters and simply say "this is the letter a", "this is the letter b", then after some time he is able to differentiate the letter a from the letter b and not only in the particular forms which were shown to him but in any other form.

Note that in this case the teacher not only does not explain how to differentiate the letters he does not know exactly how he does this himself.

A machine reproduction of precisely this process of learning was the objective of the first stage of our study.

So far, in using the term "visual image", I was working only on the basis of intuition. Let's try to give a more precise definition of this concept.

By visual image we understand an infinite number of visual perceptions, for example the result of observation of various plane figures which are notable for the following reasons: individuals who have seen only certain representatives of this number can confidently relate others which he has never seen to a particular category.

For example, a set of portraits of all the participants in some sort of meeting is not an image. A person who had not been there, when shown any of these portraits, could not when shown a new portrait say whether or not the individual shown in the new portrait took part in the meeting. But the concept "man's portrait" is an image. Although no one has ever seen portraits of all men, anyone who looks at a new portrait can with a great degree of certainty say whether or not it is a portrait of a man. Other examples of images are con-

cepts such as "the number 2", "the letter a", "a circle", etc.

When we talk of teaching a machine to recognize visual images we have in mind the following experiment. Imagine that a screen (Fig 1) is composed of a large number of photoelements on which we project some sort of image, for example the outline of the numbers 2 and 3. A current travels along a wire from each photoelement to a device which we call the machine.

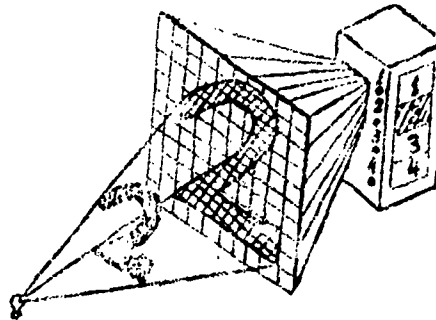


Fig. 1.

Schematic of the experiment. The object is projected on the photoelements -- receptors. Information about the state of the receptors (current from the photoelements) is fed to the machine. In the learning process a special signal (pressing of a knob on the panel for example) tells the machine to which image the projected object belongs. After learning, the machine is able to guess (and light up the corresponding panel) to which image the object belongs.

In the learning process we select several, say 20, outlines of the numbers 2 and 3 from among the infinite number of possible outlines and show them to the machine by projecting them on the photographic field. By pressing a specific knob during each demonstration we tell the machine: we showed you a 2, 3, etc. After we have shown all the selected figures the learning process is terminated and the knobs are turned off. Then we projected on the screen previously unused outlines of

2's and 3's. By use of some sort of conditioned signal (turning on a panel light for example) the machine should answer the question: "What is this".

We will be able to say that we have reproduced the learning process if in the overwhelming majority of the cases the machine gives the correct answers.

But this is still not all. We require that without any changes in the machine design or program it be able to repeat this same experiment using new images, for example we should be able by the same procedure to retrain the machine to recognize letters, geometric figures, or to train it to differentiate between portraits of men and women, etc.

THE COMPACTNESS HYPOTHESIS

In this statement of the problem it is not known ahead of time what specific images the machine is to be taught to recognize. This means that it will not be possible to store in its program a collection of signs suitable for the recognition of a given image.

We can see only two methods which might aid in the solution of this problem. Either the machine program should contain so many of the most varied signs (which are either stored ahead of time in the machine or evolved by the machine) so that this collection will be adequate for the recognition of different images or it is necessary to find and insert into the program either one or a small number of universal signs which are fundamental to the very concept of "image".

It seems to us that such a universal sign exists and that it is closely allied with the definition of the concept of "image". This sign might be formulated in the form of the hypothesis proposed by E.M. Braverman and called the "compactness hypothesis".

We will begin our explanation with a specific example. Figure 2 shows several outlines of the figures 5 and 3. If we were to ask a large group of people what each of these figures is, it is obvious that there will be complete agreement that the upper row contains 5's and the lower row 3's. As for the figures in the center row there will be disagreement and there will not be an overwhelming majority for either the 3 or the 5.

Consider the following situation. If we in some arbitrary fashion slightly modify any figure of the upper or lower row then these new figures will be easily re-



5 5 5 5
 5 5 3 3
 3 3 3 3

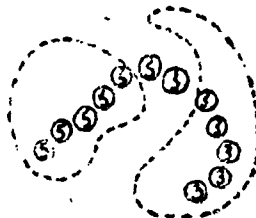


Fig. 2.

The numeral 5 is gradually transformed into the numeral 3. If we could "see" in multi-dimensional space we would see approximately the following pattern: the dashed line to the left enclosing the 5's area, the dashed line to the right enclosing the 3's area, with the small circled figures located in one or another of the dashed enclosures or between these enclosures corresponding to the figures of the upper portion of the diagram.

cognized. At the same time the extreme left 5 and the extreme right 3 in the center row will cease to be easily recognized only in the case of certain specific changes (deformation of the upper parts) and in the case of any other slight changes these figures continue to be easily recognized. The compactness hypothesis expresses this

obvious fact and considers that it is common to all images.

We shall now formulate this hypothesis in more precise terms. Consider that our screen contains only two photoelements (Fig 3,a). If we lay off along the coordinate axes the status of each of these elements we will be able to characterize the status of the photo screen by a point on a plane. If the screen consists of three elements (Fig 3,b) and is considered as a three-dimensional space along the coordinate axes of which are layed off the status of the photoelements then the point of this space characterizes the three-element screen. Similarly, a screen containing n photoelements can be plotted in the corresponding n -dimensional space so that a point of this space corresponds to every state of the screen. We call this the receptor space (Fig 3,c).

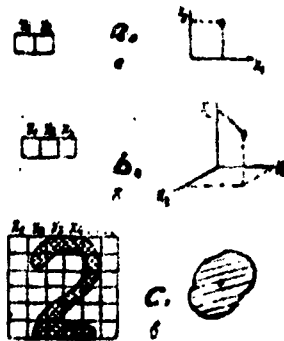


Fig. 3.

The status of two photoelements is described by a point on a plane in a two-dimensional space (a); the status of three elements by a point in three-dimensional space (b); the status of the photoelements when an image is projected on them is described by a point in multi-dimensional space (c).

Let's project some sort of image on the screen. All the elements will take on some state, that is a definite state of the screen will correspond to the given image and thus to a specific point in the receptor space.

The compactness hypothesis states that to all the

patterns making up any image in a space there corresponds an area and that for different images there are corresponding different areas which do not have common points.

From now on, when we consider such a n-dimensional image we will arbitrarily show it on a plane (see lower half of Fig 2).

Using the considerations described above we will now formulate the problem of teaching a machine to recognize visual images. In the receptor space there are really two separate areas -- one of them corresponds to the first image, the number 5 for example, and the second corresponds to the second image, the number 3. For the sake of simplicity we will from now on always be talking about teaching a machine to recognize two images. Actually there could be any number of images without requiring any change in proposed teaching algorithm. The boundaries of the areas are not known ahead of time. In the process of teaching the machine is shown some quantity of randomly selected representatives of these images, that is some quantity of random points from the first and second areas. A problem is presented to the machine: knowing only these points, to draw a separating surface so that the areas are located on opposite sides of the surface.

After completion of the teaching process, when new images, or what is the same thing new points in the receptor space, are shown the machine will be able to answer the question of which image these points correspond to. And it will be able to solve this only as a function of which side of the separating surface the points lie on.

If the separating surface is drawn properly and the areas are completely separated by it then the answer will always be correct no matter how many new patterns associated with these images are shown.

ALGORITHMS OF THE LEARNING PROCESS

Two machine algorithms were proposed for the solution of the problem of the drawing of the separating surface.

The first is the algorithm of the "random planes". Assume that in the learning process two points first appeared and the machine was told that they corresponded to different images (the number 0 and 1 appeared, for instance). Then the machine will draw a plane which is selected at random but with one limiting condition: it

must separate the points (Fig 4,a). If another point now appears and it is found that there are two points belonging to different images and which are not separated by the plane then a new random plane is drawn separating them (Fig 4,b) and this is continued throughout the learning process.

And thus at the end of the learning process there is stored in the memory of the machine a large number of randomly drawn planes (Fig 4,c). They divide the entire space into areas which contain similar points or which may contain no points at all. If on both sides of any plane (or part of a plane) there lie identical areas then this plane is not required in order to separate the areas. Therefore, after the learning process is completed in accordance with a specific algorithm we "erase" the pieces of the planes on both sides of which similar points are located or the pieces of planes on one side of which are located points used in the learning while no points at all are found on the other side. As a result, a surface is formed which consists of pieces of planes (Fig 4,d). The surface thus formed is taken to be a separation surface. In the future when a new pattern is shown -- new points appear -- the machine will answer the question of what is this form on the basis of which side of the surface the point lies on.

As a rule, however, the constructed surface will not separate the areas precisely (as shown in Fig 4,d) and as a result the machine will err.

It is possible to drastically reduce the number of errors in the following manner. Imagine that several machines are being subjected to the learning process at the same time. Since the planes are drawn in a random fashion each machine will draw different separating surfaces which means that each machine will err on different patterns. Now if in the recognition of an unknown object the decision is made on the basis of a majority vote of all the machines then it is clear that the probability of correct answers will increase sharply. This method has been given the name of "paralleled alternatives".

Now consider the second algorithm which is based on the construction of potential surfaces. In the learning process when any sort of point is shown to the machine it constructs a function which reaches a maximum at the given point and, while remaining positive, reduces on departure from the point in any direction. As the machine is shown more points belonging to a particular area

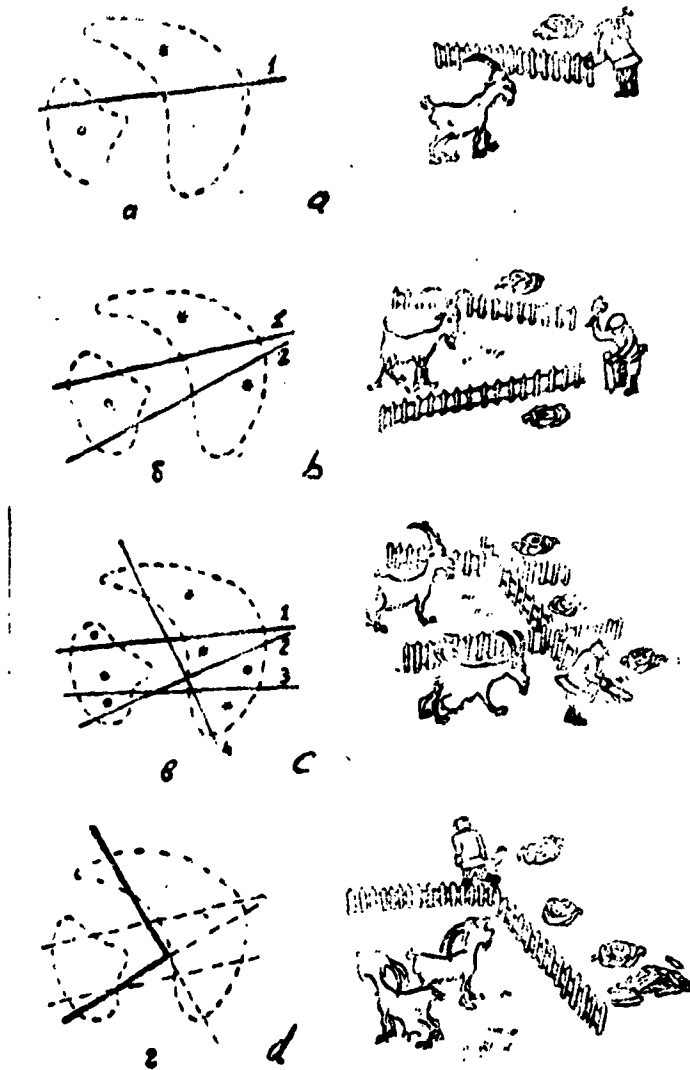


Fig. 4.

The "random planes" algorithm: a - in the receptor space appear two points corresponding to different images and they are separated by the plane; b - the third point falls on the same side of the first plane as the point belonging to the other image and therefore they are separated by a new plane, and so on; c - the learning is completed and all the points are separated from all the asterisks by planes; d - the superfluous planes and pieces of planes are removed leaving only the separating surface.

it constructs these functions for each of these points and then adds them. As a result the machine constructs a potential surface which has a maximum value ("hump") at the points of the area and reduces sharply away from these points.

In a similar way the machine constructs a function having a "hump" at the point belonging to the second area.

The learning process is terminated by the construction of this potential function. In the recognition of "unknown" objects, that is when new points appear in the receptor space, the machine relates them to that area whose "potential" is greatest at that point. In the case of the second algorithm the separating surface is the projection of the line of intersection of the constructed potential surfaces.

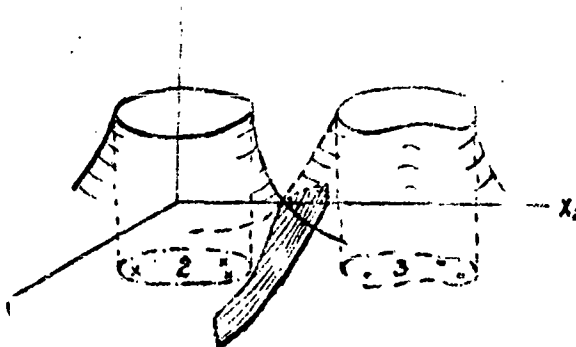


Fig. 5.

Algorithm of potential surfaces. During the learning process there are as many functions constructed in the receptor space as there are images which the machine is to be taught to recognize. Each function corresponds to a single image, it has a "hump" over the area belonging to this image, and decreases away from that area. In the recognition of a new object the machine takes into consideration the value of each function at the point corresponding to this object and relates the unknown object to that image whose function has the greatest value at this point.

EXPERIMENTS

In order to verify the compactness hypothesis and the two algorithms which we have described experiments were conducted on a general-purpose digital computer.

In the experiments using the first algorithm we prepared 160 outlines each of five figures: 0, 1, 2, 3, and 5, or a total of 800 outlines of numbers. We chose 40 shapes of each numeral to use for the learning process (the other 600 were used for the subsequent testing of the machine).

In each case the number of correct answers given by the machine ranged as high as 89% and in the case of the "paralleled" procedure ran as high as 98%. At the end of the learning process from 1,500 to 3,000 of the binary memory units of the machine were occupied.

In the series of experiments using the second algorithm we used either five or all ten of the digits simultaneously. In this case we used only a quarter as many samples of each number, 10, in the learning process. The machine gave 100% correct answers using this algorithm to differentiate the five digits (out of 750 interrogations). During the tests with teaching the machine to recognize simultaneously all ten digits the figure of 85% correct answers was obtained in 1,400 interrogations. It is now clear to us that it is necessary to improve the algorithm in order to get a figure of 100% correct answers in this case.

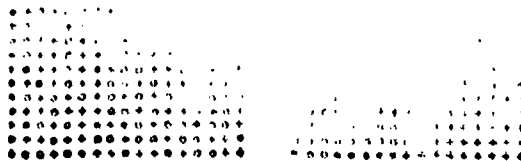
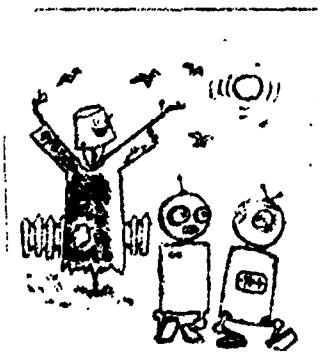


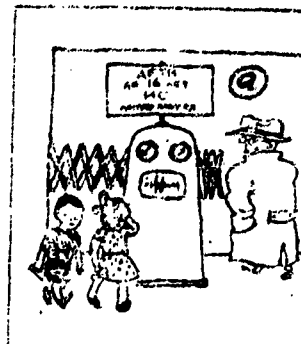
Fig. 6.

These diagrams clarify the concept of compactness. On the left the crosses and zeros are not placed compactly and if you try to separate them with a single curve you will find it very difficult. On the right the crosses and zeros are positioned compactly and it is very easy to draw a separating curve.

At the present time experiments are being conducted on the use of the algorithms described to teach machines to recognize more complex images, for example to recognize simultaneously all the letters of the Russian alphabet written in script, to recognize letters set in differing typographical characters, to operate in the presence of "noise" (smears, blots, random shifts of the letters, etc.), to differentiate male and female portraits shown both front-face and profile, etc. Although these experiments have not been completed the successful results of the tests with the numerals is encouraging and we believe that we will be able to achieve good results in the more complex cases.



This fellow must really love nature



Don't cry, we'll soon be adults, too
 ⓐ - children under 16 not admitted.

In conclusion I should like to mention the physiological and psychological experiments which are being conducted during this program under the direction of Professor S.N. Braynes in the USSR Academy of Medical Sciences. These tests have only recently been initiated and it is still too early to discuss their results but the preliminary data are very interesting.

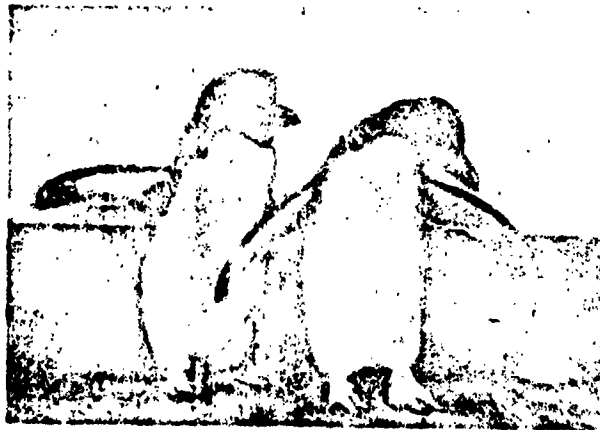
The psychological tests are being conducted with two series of blots. Each series contains 150 blots specially "drawn" for these tests by machine so that all the blots of a given series belong in the receptor space of one area and all the blots of the second series belong in the other area.

During the experiments intermixed blots of the two series are shown rapidly to children and they are asked

to separate them into the two groups. Since the blots are devoid of any sort of meaning and the outlines are random they can not be differentiated by means of any simple, immediately apparent features. Successful completion of this experiment would be strong proof in favor of the concept that living organisms recognize such varieties of simple, "meaningless" images simply as a result of their compactness.

The physiological tests, which are akin to the psychological tests just described, are being conducted using rats and monkeys. Attempts are being made to develop in the animals reflexes to the appearance of blots from different series and from the reaction of the animal to judge to what extent he is able to differentiate blots which are similar or differentiated in only one feature -- they correspond to points forming different areas in the receptor space.

At present it is difficult to judge just what all this study will mean for biology and physiology. But for the field of engineering it is clearly of considerable importance since it permits us to hope for significant expansion in the capabilities of machines. Today, if machines are to perform different functions it is necessary to construct (program) them differently. But now we have some hope that the machines can be taught differently if trained differently rather than built differently. It is difficult to overestimate the possibilities which this opens up in technology.



These penguins were brought from Antarctica in May of 1962 and are now in the laboratory of Professor S.N. Braynes. They are to be used in experiments on image recognition.

10484

CSO: 1872-S

- END -