# ARisTotle: A Neural Net Architecture for the Auto-classification, Grouping and Recognition of Partially Distorted Binary Images.

by
D. Boulanger and D. Poussart
Laboratoire de Vision et Systèmes Numériques
Département de génie électrique
Université Laval
Québec, Canada
**e-mail:** boulanger@gel.ulaval.ca
poussart@gel.ulaval.ca

## Abstract

In this paper we present a classification network for the recognition and hierarchical grouping of binary patterns based on a ART 1 neural classifier. The network is first presented with a series of primitive patterns which are memorized. After each presentation of a complex pattern the network automatically selects the classes of the previously learned primitives that match the parts of the image shown. For each primitive selected, the network spatially isolates its couterpart in the image and computes the Hamming distance between the two parts. After recognizing all the parts of the complex image, a grouping of these constituant classes is made to a new class for the complex image. This hierarchical representation is use afterwords for the recognition of partially distorted complex patterns.

The network was trained for recognizing handwriten numerals. After the presentation of a few primitives and only one example of the numbers from 0 to 9, the network was able to identify many distorted and noisy patterns of numbers.

**Keywords:** Neural net, hierarchical, classification,ART network, image processing, character recognition.

## I Introduction

Aristotle thought that things in the universe could be classified in five groups. These groups, structured in an ascending hierarchical order, extended from minerals to pure thought. Man, a class by itself, was the highest form of being on this planet and was placed just below pure thought. A class was thought superior to an other if it possessed the characteristics of the lower classes and some of its own [1].

It was evident to Aristotle that man's knowledge was also driven by this hierarchical structuralism, as in nature. Man's knowledge is the result of his experiences acquired through his senses.

This philosophy of man's knowledge structure still has an influence on today's comprehension of the complex structure of our visual system. In spite of the complex world surrounding us, our visual system, from the retina to the cortex, is able to capture this world with a certain ease.

Primary information is first extracted from a complex scene, and is then classified at different levels of representation helping us afterwords, to recognize things of the same nature [14].

When analysing a complex scene, our visual system seems to follow a hierarchical search from the most dominant features to the less. These features, like edges, color, texture and others, excite in parallel the different cognitives levels and induce images or models of things that are compared with the inputs. This hierarchical search is generally non-oriented, meaning that the initiative of this search originated from external stimulation and not from internal thinking like the search of a particular object in a complex scene.

In our approach we have tried to keep these principles in perspective. The parallel algorithm described in this paper is able to make a hierarchical classification of binary patterns. When a complex pattern is presented to the net, a fast and direct search of its constituent parts is made from the previously learned ones. The classes for these recognized parts are then grouped together with the class for the complex image resulting in an ascendingly complex representation. These complex groupings, now part of the knowledge base, are then used subsequently for the identification of still more complex patterns.

This system, mainly based on an ART 1 neural classifier, is capable of identifing partially distorted patterns by means of this hierarchical representation. In the first part of this paper the ART 1 system is briefly described with an emphasis on the emergent properties of this neural architecture. The proposed system, named ARisTotle, is then described and compared with an ART network. Finally the properties of ARisTotle are demonstrated by applying it to the task of handwriting recognition.

## II ART 1 system

While developping their adaptative resonance theory, Carpenter and Grossberg [3,8,9] described a number of classification models. The ART 1 system [2] implements a massively parallel algorithm for the non-supervised classification of binary patterns. The network is made of two layers of units or neurons. Units of the first layer called F1, serve as input as well as comparison units.These input units are tied to the classification units by a series of weighted links wich adapt slowly in order to memorize the input patterns. The classification units, call F2, select the unit receiving the maximum excitation from F1 layer ("winner-take-all"). The network also has a reset unit for the validation of the selected class and a gain control unit for switching the F1 units from the input mode to the comparison mode.

The search process, illustrated by Figure 1, is as follows:

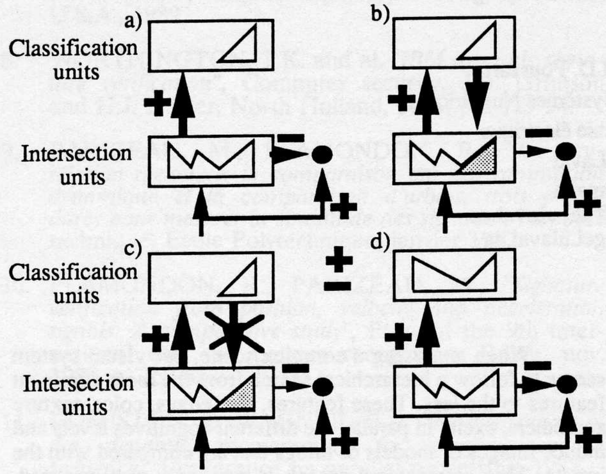



Figure 1 Search process in an ART 1 system: a) The network first selects a class through the bottom-up links b) An image of the selected class is then fed back to the input layer. c) If the relative difference between the input pattern and the class is greater than a certain threshold, the selected unit is reset. d) The process continues until the selected class matches the input.

The network begins by comparing the inputs in parallel with all the memorized models using the bottom up interconnections between the two layers. This generates a series of excitations to the classification units. The unit receiving the highest excitation is then enhanced and selected. This selected unit then generates an image of the memorized model through the top-down links to the first layer. The intersection of the generated model and the input is then computed at the first layer using the gain control unit. If the relative difference between the intersection and the input is lower than a certain factor called the vigilance factor, the selected classification unit is deactivated temporarily by the reset unit. The condition for reset is:

$$\frac{|X \cap I|}{|I|} > \rho \quad \text{where } 0 \le \rho \le 1$$

In other words if the euclidian distance between the presented image and the selected model is too large, forget this selected class and try another one.

After the reset, the process is re-activated and another classification unit is selected and validated. This continues until a match is found which gives a relative difference higher than the vigilance factor. If there is no reset, than the net is said to be in resonance. If the input is new to the net it will automatically classify it as a new unit.

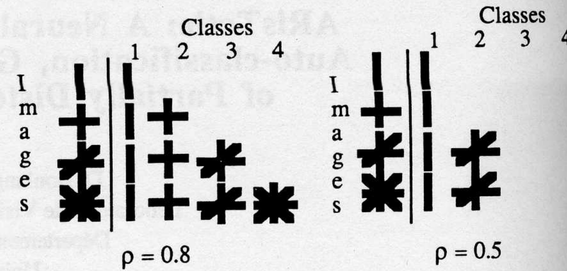



Figure 2 Influence of the vigilance factor. The number of different classes learned rises with the vigilance factor while its noise sensitivity diminishes.

Figure 2 illustrates the influence of the vigilance factor on the number of new classes generated. If the vigilance factor rises, so does the number of classes generated but its sensitivity to noise also rises. If this factor is lowered, its sensitivity to noise diminishes but so does the number of discriminating classes.

Every time the net is in resonance, the bottom-up and the top-down links between the two layers are gradually modified to adapt to the variations between the input and the model.

The values of the weights for these links are chosen in such a way that inputs of the same nature will have a direct access to this class without searching through all the others.

## III ARisTotle: A hierarchical binary classifier

One of the problems with an ART classifier is that it cannot identify the constituent parts of a complex image. This global approach does not provide a good symbolic representation needed for recognition. The proposed system has the same properties as the ART system, such as direct access, but it can also cope with the parts of a complex image by using a spatial delimitation network. Every object in a complex image generates, as in an ART system, a series of excitations to the classification units which are selected one after another to be compared with the whole image or parts of it. With the spatial delimitation network, the selected models can be classified as being a part of the presented image or as the whole.

After every part is recognized by the net, the whole image is classified with a different unit. A grouping of this last class with all the sub-classes found previously, is then made. This symbolic representation of complex images is used later on for the recognition of partially distorted images of the same nature.

### III.1 General description of ARisTotle network

The net has six main sections. A spatial delimitation network which is able to focus on parts of the image. A classification network which memorizes the patterns. Another section is used to temporarily memorize and make hierarchical groupings of the validated classes. Finally, control units are used to validate the selected classes while others make an asynchronous update of the net possible.

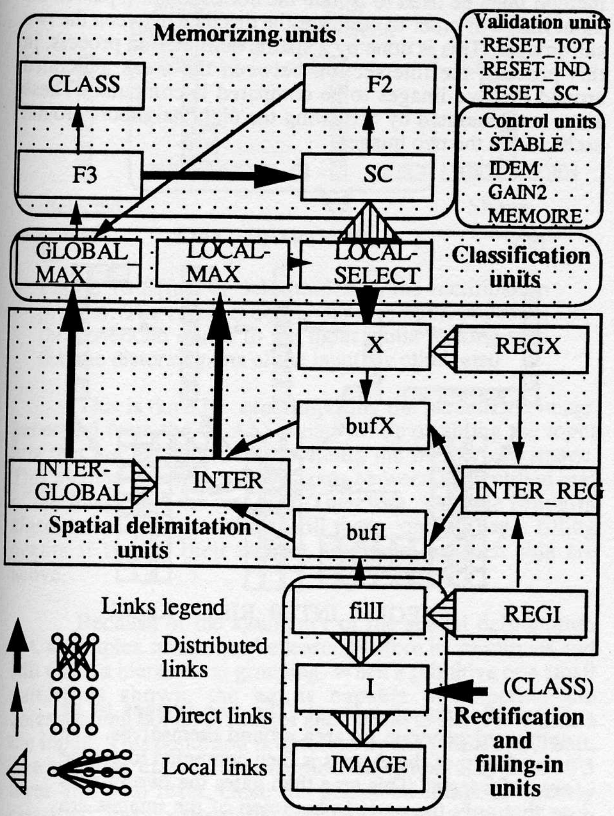



Figure 3 The ARisTotle network. The network consist of made of six sections: the control units, the validation units, the spatial delimitation units, the classification units, the memorizing units, the rectification and filling-in units.

### III.1.1 Class selection and image focussing

ARisTotle has two types of classification units. First, the units called LOCAL_MAX are used to classify the images at nine shifted positions around the input image.

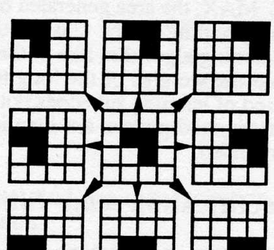


Figure 4 Nine positions memorized by LOCAL_MAX. Each new pattern is memorized at LOCAL_MAX by shifting by one bit every bit in the input pattern in each direction.

This type of representation will enable us to recognize distorted patterns. Every unit in each group of nine are linked to one another in such a way that only one unit of each group is activated at any one time. When a shifted pattern is presented to these units, only one is actived, disabling all other units in the group.

The second type of classification units, called GLOBAL_MAX, are used to memorize the area generated by the successive shifting of the input pattern.

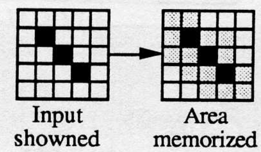



Figure 5 Area memorized by GLOBAL_MAX. The GLOBAL_MAX classification units memorize the area generated by the succesive shifting of the input.

Each unit represents one pattern. Units are linked together in such a way as to limit activation to a single cell. The selected class will be the one corresponding most closely to the input, like in the ART system. Even though the input pattern is distorted, the corresponding class is still activated because it falls in the memorized area.

The network sequence of execution is as follows. After the presentation and learning of the primitives and prototypes of complex images, the input units process a complex image made of those previously learned patterns. This image generates a series of excitations through the links of the long term memory to the classification units LOCAL_MAX and GLOBAL_MAX. The unit in GLOBAL_MAX receiving the highest excitation is activated while all others are disabled. One unit in each group of nine from LOCAL_MAX is also activated.

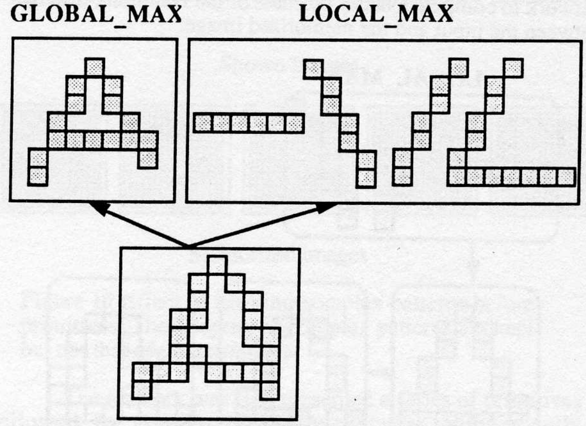



Figure 6 Class selection . At GLOBAL_MAX only one complex class is selected while at LOCAL_MAX one class in every group is selected.

The active unit GLOBAL_MAX generates through F3 a series of sub-classes associated with the complex selected class. These sub-classes in turn select the corresponding groups at LOCAL_MAX.

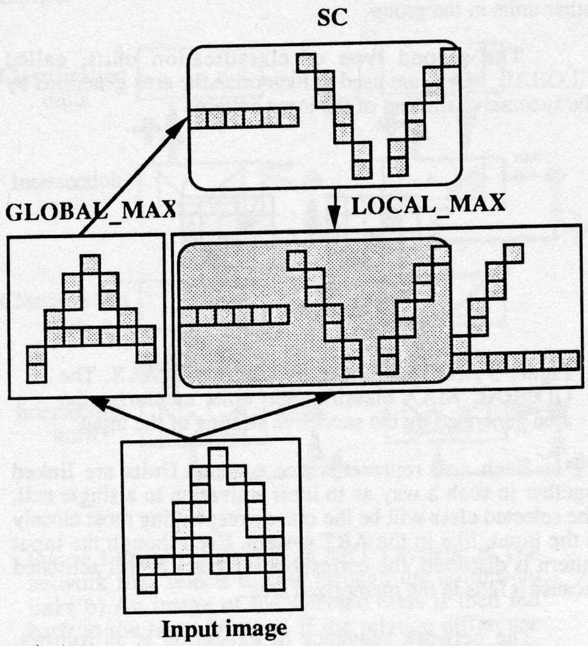



Figure 7 Generation of the associated sub-classes. The selected class at GLOBAL_MAX activates the associated sub-classes which help to select the corresponding shifted sub-images at LOCAL_MAX.

As in the ART system, an image is then generated within the X layer. This composite image is obtained by superimposing the selected primitives shifted in a way that best matches the input image. This representation enables the network to compute a better estimate of the Hamming distance between the input and the memorized image.

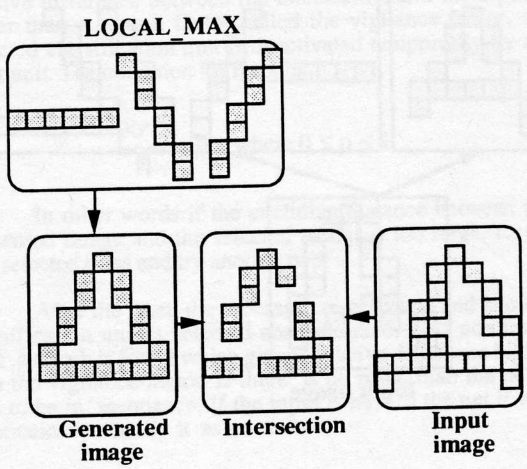



Figure 8 Generation of the selected image. The selected sub-classes from LOCAL_MAX create a distorted image that will be compare with the input for validation.

If the selected image is only a part of the input, a method must be used to isolate the corresponding part in the input image in order to compare the distance between the two sub-images. This is done by a spatial delimitation process. In this process, the intersection between the areas generated around the two images to be compared is computed. These areas are generated by activating the eight neighbors around each unit in the two images.

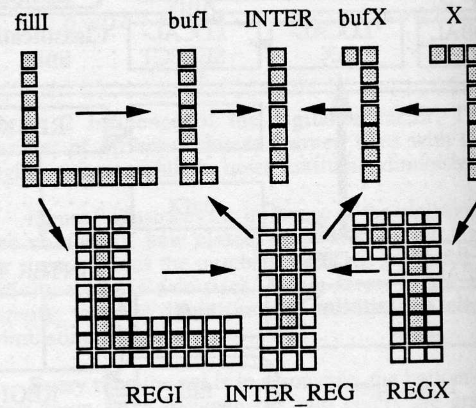



Figure 9 Area of interest. The two images to be compared generate an area around themselves. The intersection of these areas is then computed to get the area of interest. This area then gates the two images so that only the parts in common of the images are compared.

The intersection gives us the area of interest. This area is then used to gate the bits of the two images which are to be compared.

A real comparison of the relative distances of these two gated images can then be computed from their intersection. If the relative distance between them is lower than a certain threshold, the selected class and sub-classes are reset, otherwise it is temporarily memorized in the F2 layer, and then reset. The process is then re-activated and continues until the whole image is classified in a previously learned class or in a new one. The links of the long term memory are then modified to memorize the image in its nine shifted positions at LOCAL_MAX, the area generated by this shifting and the hierarchical grouping of the class with its selected sub-classes. A fast learning algorithm like in the ART system was used to modified the weights of the long term memory. Unfortunatly, this kind of learning rule does not garantee that the input and the memorized pattern are identical because a small distorsion is tolerated.

### III.1.2 Rectification and filling-in

In order to insure a constant and undeformed representation of the memorized patterns, the input image must be rectified in the same way as the memorized one. Although the input is only a part of the selected image the input must be modified so that it matches the corresponding part of the complex image. This rectification is done by superimposing the memorized image on the input by making sure not to keep the active bits from the distorted image.

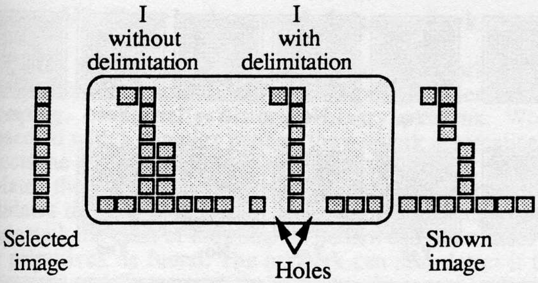



Figure 10 Effect of delimitation in the rectification process. The rectification is done by superimposing the generated image to the input while making sure that the distorted parts of the input are eliminated.

This is done by superimposing the memorized image generated from the CLASS units, while disabling the input bits that are in the area surrounding the memorized image. This method sometimes leaves gaps between the rectified part of the image and the rest of it. Therefore a simple filling-in algorithm is used in order to fill these one bit gaps: filling occurs if two of their nearest neighbors on each side are active.

Because of the symmetry of the spatial delimitation net, a complex image can be learned before its primitives and still make a hierarchical grouping. When a primitive or a small pattern is shown, the net is capable of selecting the coresponding complex images and is able to tell if it includes the input. This condition is detected by the RESET_SC unit. The complex class is then memorized temporarily at the F3 layer and then reset. Once the presented image is entirely classified, links between the complex classes associated with the input image are made with the constituant classes of the input.

## IV Control units and simulations

The network has been simulated using the "Rochester Connectionnist Simulator" (RCS) wich has been developped at the University of Rochester [7].

A simplified version of the membrane equation was used in the simulation in order to compute the activation levels of each units. The activation for each unit is given by:

$$O_k = f\left( J_k^+ - J_k^- \right)$$

where

$$f(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

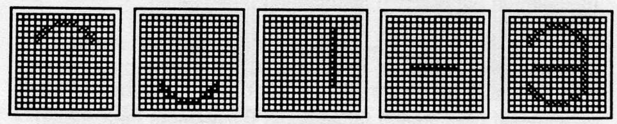
where $J_k^+$ is the total positive excitation and $J_k^-$ the total negative excitation.

Sometimes the activity of some units must vary more slowly than in others. This is achieved by adjusting the time constants of the membrane equation at different levels to synchronize the order of execution in an asynchronous computation of the net. For further simplification, control units were used instead of adjusting these time constants. Control units are activated only by specific states in the simulations and they are used to synchronize the activity of certain layers at specific times in the simulations.

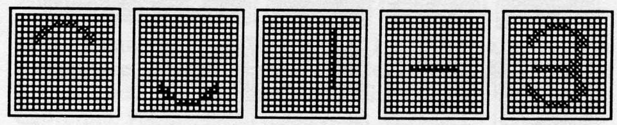
## V Handwriten number recognition

For illustrative purposes, the network was trained to recognize handwriten numbers from 0 to 9.

Because of the rectification process, the order of presentation of the patterns is important. In the case where the primitives are presented before the complex pattern, the integrity of the primitives are preserved but not that of the complex image.



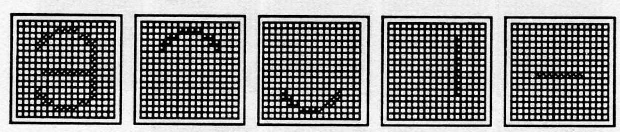
Shown images



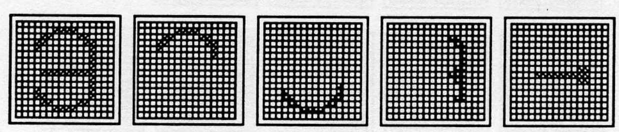
Memorized images

Figure 16 Effect of learning primitives before complex patterns. The integrity of primitives are kept but not that of the complex image.

However, if we present the complex image before the primitives we preserve the integrity of the complex image but not that of the primitives.



Shown images



Memorized images

Figure 17 Effect of learning complex patterns before primitives. The integrity of complex patterns are kept but not that of the primitives.

The network was first presented a series of primitives followed by one prototype.The network automatically decomposed the numbers as follows:

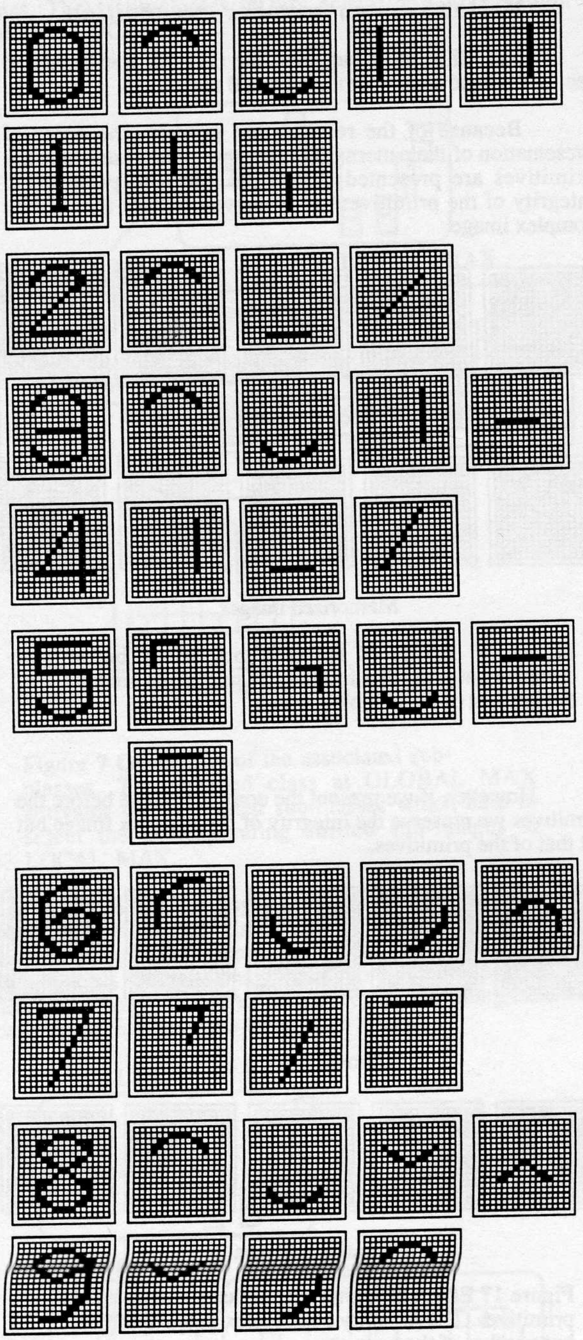


Figure 18 Numbers and associated primitives: After learning a few primitives a prototype of each number was presented. The network then decomposed the numbers into the primitives.

The network was then shown a few partially distorted numbers, some with noise. The network then proceeded with the recognition and classification of the shown patterns.

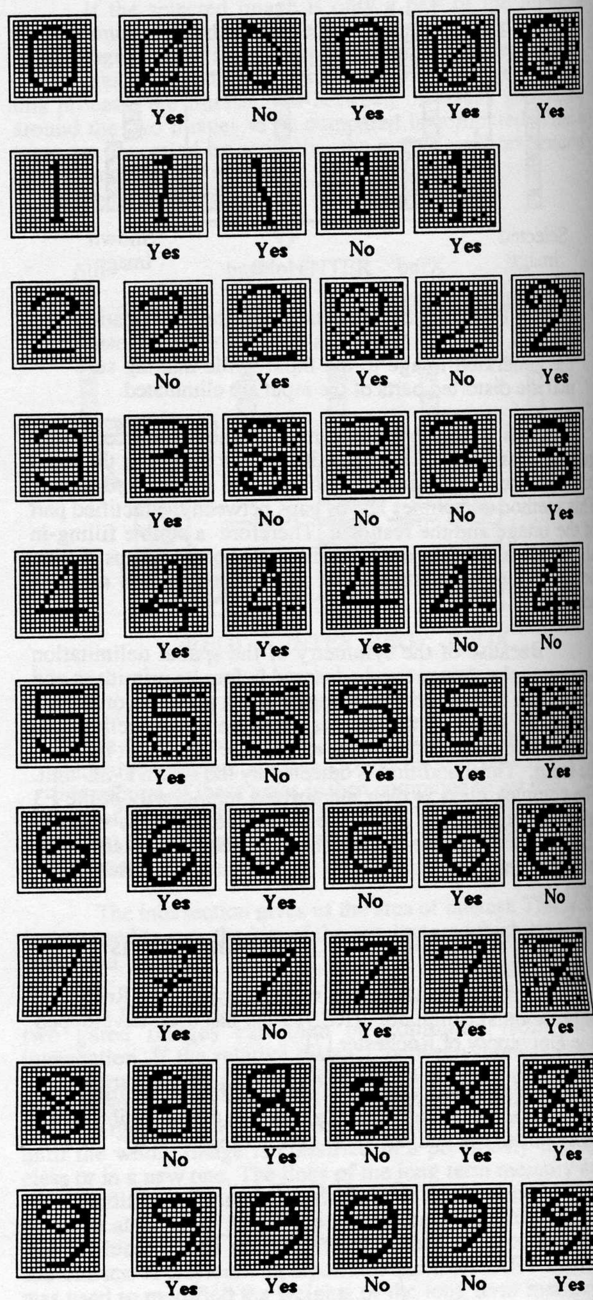



Figure 19 Recognition of distorted numbers. After the learning of one prototype the network was able to recognize distorted and noisy images.

Most of the time, the first class activated corresponded to the number shown. However, similar numbers sometimes excited the wrong class at first, but were invalidated by the network. The proper class was then selected and validated.

Depending on the types of primitives used the recognition rate is different. In general a segmentation with more independent primitives provides a better recognition rate than one with less primitives.

## VI Conclusion and future work

In this paper we have presented a network for the recognition, the auto-classification and the hierarchical grouping, of partially distorted binary patterns. When presented with a complex pattern, the network automatically selects the constituant primitives previously learned, spatially isolates their corresponding parts on the input pattern and validates the selected class. A hierarchical grouping is then made with the class of the complex pattern and the classes of the primitives is found. The network can also detect if the input image is a part of the complex images previously learned. It is also able to recognize partially distorted patterns. The activity of every unit is described by a simplified version of the membrane equation.

The network was tested for the recognition of handwriten numbers from 0 to 9. After the presentation of only one prototype of each number the network was able to recognize a series of distorted and noisy numbers.

Although the present work is limited in its applicability, we ought to remember it's principal properties: the focussing process, the direct access to a class and the rectification and adjustment of the input during its analysis.

Future work will emphasize these properties for the classification of images (2-D and 3-D) represented by frames. This will permit the mixing of different types of information, giving a more elaborate model of learned objects.

## Aknowledgements

This work was supported in part by grant A5274 of the Natural Sciences and Engineering Council of Canada (NSERC), grant 89EQ2830 of FCAR (Province of Québec) and funding made available through the network of centers of excellence of the *Institute of Robotics and Intelligent System*. A special thanks to John Laurent for his help in revising the manuscript.

## References

[1] **Adler M.J.**, "Aristotle for Everybody", MacMillan publishing Co inc.,New-York,1978, pp 4-167.

[2] **Carpenter G.A., Grossberg S.**, " A Massively Parallel Architecture for a Self-organizing Neural Pattern Recognition Machine", Computer Vision Graphics and Image Processing, 1987, pp 54-115.

[3] **Carpenter G.A.**, Grossberg S., "The ART of Adaptative Pattern Recognition by a Self Organizing Neural Network", Computer, mars 1988, pp 77-89.

[4] **Cohen M.A., Grossberg S.**, "Some Global Properties of Binocular Resonances: Disparity Matching, Filling-in and Figure-ground Synthesis", Mathematical Department , Boston University, Mass., 1983, pp 1-70.

[5] **Fukushima K.**, "Neural Network Model for Selective Attention in Visual Pattern Recognition and Associative Recall", Applied Optics, Dec 87, pp 4985-4991.

[6] **Fukushima K.**,"A Neural Network for Visual Pattern Recognition", Computer, mars 1988, pp 65-75.

[7] **Goddard N.H., Lynne K.J.**, Mintz T., "Rochester Connectionnist Simulator", Rochester University, New-York, 1988, pp 1-181.

[8] **Grossberg S.**, "Adaptative Pattern Classification, an Universal Recoding: I. Parallel Development and Coding of Neural Feature Detectors", Biological cybernetics, Springer-Verlag, 1976, pp 121 -134.

[9] **Grossberg S.**, "Adaptative Pattern Classification, an Universal Recoding: II. Feedback, Expectation, Olfaction, Illusions", Biological cybernetics, Springer-Verlag, 1976, pp 187-202.

[10] **Man D., Wishihara H.K.**,"Visual Information Processing: Artificial Intelligence and the Sensorium of Sight",Reading in computer vision, edited by Fischler M.A. Firschein U., pp 616-637.

[11] Manuel du Cours: Principes de la Logique, Département de Philosophie, Québec 1987, pp 88-91.

[12] **Ryan T.W., Winter C.L.** ,"Variations on Adaptative Resonance", Proceedings of IEEE First International Conference on Neural Networks, San Diego, Californie, vol II, pp 767-775.

[13] **Ryan T.W., Winter C.L.**, Turner, C.J., "Dynamic Control of an Artificial Neural System: the Property Inheritance Network", Applied Optics, Dec 87, pp 4961-4971.

[14] **Ullmann S.**, "Visual Routines", Reading in Computer Vision, edited by Fischler M.A. Firschein U., pp 189-328.