DIRECTED SPREADING ACTIVATION IN MULTIPLE LAYERS FOR LOW-LEVEL FEATURE EXTRACTION

A.Arul Valan and B.Yegnanarayana,

Department of Computer Science and Engineering,

Indian Institute of Technology, Madras, INDIA.

ABSTRACT

Spreading activation neural networks have been proposed in literature. This paper proposes a directed spreading activation neural network model which performs a large number of early vision tasks. It is shown how directed two-dimensional(2D) diffusion followed by detection of local maxima can effectively perform feature extraction, feature centroid determination and feature clustering all on multiple scales in a purely data-driven manner. The feature map, which is the result of this directed spreading activation process can be used in learning and recognition of 2D object shapes from their binary patterns invariant to affine transformations.

I. INTRODUCTION

Visual pattern recognition involved in reading written/printed characters or distinguishing shapes is easily accomplished by human beings, but when it is attempted to design information processors that can do the same thing it presents significant difficulties. Since human beings do this task effortlessly a number of researchers have attempted to model the visual neural system of the brain. Various models of neural systems are reported in the literature for the problem of visual pattern recognition. It is now a widely-held view that visual perception is based on two interrelated processes[RYBAK 91]: parallel processing of visual information carried out automatically by mechanisms determined by neuronal organization of the retina, lateral geniculate nucleus, and visual cortex; and sequential processing related to image recognition mechanisms that are controlled by attention. In the first process, detector properties of single neurons and local neuron nets are of primary importance. In the second eye movements are considered to be of importance. Through these movements, the most informative parts of the image are sequentially projected onto the fovea for finer processing.

Experimental research on vertebrate visual systems has revealed various structural and functional mechanisms to support this theory. [HUBEL 77] reported that in the visual cortical areas, there are multilayer organizations of largely identical cells and these layers are further divided into columns. The complexity of cells varies across layers, indicating functionally distinct roles. An area separate from the visual cortex, the superior colliculus, seems to be concerned with directing eye movements[BREITMEYER 86]. The retina has a high-resolution fovea at its center to which any area in the field of vision can be directed by physical movements. The attention seems to be concentrated near the fixation point. It appears from human eye-motion studies that there are other advantages to integrating eye movements into a computer vision system, such as a mechanism to obtain information about spatial relations and translational invariance. This active processing approach to visual processing simplifies the spatial cognition problem.

Evidence for rapid diffusion like phenomena can be found in the brightness and color domains of stabilized image experiments. Compelling evidence is provided by [YARBUS 67] experiments, in which color from the surrounding rapidly fills regions in which stabilized images have faded. This type of effect has motivated [Cohen 84] to postulate a diffusion layer ("filling-in-syncytium") as an essential component of their feature contour system.

[SEIBERT 89] proposed that diffusion enhancement can be used as a low-level computational model in building a neural network vision system. This model is used for learning and recognizing two-dimensional (2D) binary patterns invariant to location, orientation, and scale. In their model, the processing is divided into layers, each of which may encompass many levels of neuron- like processing cells. The activation is spread from the low- level feature detectors to the adjacent regions, where activity interaction occurs. In this way corners along edges, high curvature points and line-end features are located from low-level in a completely data-driven manner. The result at this point is a retinotopic map of features (feature map). Depending on the time elapsed the spreading activation layers will localize centroids in small local areas or larger, more global areas. The centroids may be used as fixation cues to drive rapid eye/camera movements (saccades), spatial relation cues for learning and recognizing hierarchies of components or invariant featural representation of objects.

Spreading activation is essentially an averaging process. When the input pattern is directly presented to the spreading activation layers, as time progresses the activation values of the individual neurons reflect the averaging process which takes place over two- dimensional space. This kind of averaging is unconstrained i.e., there is no limiting factor for the spreading of activation in both time and space. The local maxima formed as time progresses, give rise to various features and feature clusters. However, as there is no constraint in the spreading, it is very difficult to determine *a priori* when to stop the spreading process and identify the feature or the feature cluster. When spreading is not stopped at appropriate time, the peaks which are formed during the spreading slowly drift away towards the global centroid. To overcome this problem the feature map instead of the direct input pattern, is considered as input for spreading. [SEIBERT 89] has proposed location of quasi-static points during the spreading activation process as a temporal event for determination of feature clusters. This quasi-static points method cannot be adopted to feature extraction directly as the feature maxima tend to move faster towards the global centroid. So the spreading activation is used directly on the input pattern for locating only the corners. The feature map is formed only using the corner locations and is used for subsequent processing. But the lines, curves of different curvatures and edge termination points which are missed are very useful and significant for the higher stages of invariant pattern recognition system. Moreover when the eye/camera movement is used to identify the features located at the maxima points, the lines and contour termination points will be missed.

The drawback of the current model's inability to detect the lowlevel features like line segments, corners, curves and contour termination points correctly as part of the lowlevel feature extraction can be attributed to mainly the unconstrained nature of spreading both temporally and spatially. This paper proposes directed spreading model which constrains the spreading spatially. The spreading takes place in specific predetermined directions and the directions specified by the input pattern. The directed spreading activation model detects the lines of different lengths, curves of different curvatures and edge termination points in a purely data-driven manner.

The non-stationary nature of the feature maxima is mainly due to the lateral influence of the adjacent feature maxima. The line peaks and the peaks of the corners may be considered as complementary features. Since the spreading is unconstrained these complementary feature peaks spread quickly and become nonstationary. To avoid this lateral influence it is necessary to separate these complementary features. In this directed spreading activation model there are two surfaces which work parallely and locate complementary features. One layer of neurons are sensitive to lines of different orientations and the other layer of neurons are sensitive to curves of different curvatures and contour terminations.

Section II introduces the directed spreading activation model and section III shows how directed spreading activation can be used for feature extraction, feature clustering and feature centroid detection. Section IV presents a neural network model for directed spreading activation.

II. DIRECTED SPREADING ACTIVATION

Let us consider a region R and an activation function A(R) defined over it at an initial time t_0. The function A(R) is a binary valued function at t_0, either A_{sat} or 0, corresponding to locations where the pixels are 'on' in the input visual pattern. The activation can diffuse locally through the region either uniformly or nonuniformly according to the classical diffusion equation:

$$\frac{dA}{dt} = \delta\,.\,[\,k(R)\,.\,\delta A(R)\,] \quad ---- \quad (1)$$

where k(R) accounts for the density and conductivity of the region. If k(R) is a constant then the spreading is uniform throughout the region. When k(R) is a function of direction then the activation spreading takes place in specified directions. If the total activation is held constant, then the locations with initial activation A_{sat} begin to lose activation, while adjacent locations begin to gain activation. Due to superposition, areas near activation-rich locations gain activation more quickly than areas far from the activation-rich locations. Activity spreads as the time progresses from t_0 until a global activity maximum emerges indicating the geometric centroid of the features. At an intermediate time various local maxima can be detected.

Maxima detection is a problem which is well addressed in both calculus and neural networks paradigms. The activation distribution in the diffusion layer defines a surface over a 2D plane. Extrema of activity are found in areas of positive Gaussian curvature of the surface. [GROSSBERG 73] has shown that maxima can be computed in a neural network by self-activation and competition. Using lateral inhibition, each element suppresses all others with a strength proportional to its activation, while feeding back excitatory activation to itself. This is accomplished using an on-center/off-surround recurrent receptive field for each element. Among other emergent properties, this type of network enhances the contrast[GROSSBERG 73] of the activity distribution, or in the extreme case leaves only the maximally activated element on.

III. FEATURE DETECTION BY DIRECTED SPREADING OF ACTIVATION

Unconstrained spreading activation followed by maxima detection can be used to detect features like corners (high-curvature points along contours) and contour intersections. But this fails to detect the contour terminations and straight lines. This is due to the inherent nature of k(R), the conductivity function which is constant throughout the region. If k(R) depends on the direction, ie. if the spreading takes place in specific directions, this can lead to detection of lines as features and the centroid of the lines are also located.

In the directed spreading activation system proposed in this paper there are three layers(Fig. 1) each with different characteristic k(R). The first layer L1 has k(R) defined for specific directions and spreading takes place only in these directions. Hence it locates the centroids of the lines. The second layer receives its input from first layer and the input binary pattern. In the second layer the spreading activation takes place in the direction specified by the activation values of the adjacent neurons. So the conductivity function k(R) of the region is directed purely by the data. This second layer detects curve centroids of all curvatures and contour terminations. Since the spreading in these two layers is spatially constrained there is no lateral influence between peaks, hence these peaks are always stationary and the movement is restricted to the directions specified within a layer. In the third layer the k(R) is kept constant and hence it behaves like the normal spreading activation layer reported in the literature. Third layer detects the feature clusters by locating the quasi-static points. These three layers along with their maxima detectors together locate centroids of lines, curves, corners and contour terminations in a purely data-driven manner which can be used for eye/camera movement.

IV. MULTILAYER NEURAL MODEL FOR DIRECTED SPREADING ACTIVATION

The configuration of the directed spreading activation layers is shown in Fig.1. The first layer L1, consists of two dimensional array of hypercolumns. These hypercolumns receive their input from the input binary pattern. Each hypercolumn consists of a number of directional detector neurons as shown in Fig.2. All the directional detector neurons are totally connected and these links have a small negative value. Hence when the input is presented each hypercolumn act like a "winner-take-all" network as shown in Fig.3. All the directional detectors belonging to a hypercolumn receive their input from a fixed window of the input pattern. Adjacent hypercolumns receive their input from a overlapping windows.

The general structure of the directional detectors is essentially the same as that of the S-cells of Neocognitron proposed by [FUKUSHIMA 82]. Each directional detector has two types of cells, excitatory cells (ECs) and the inhibitory cells (ICs) that occur in pairs. Each pair receives the same input set. The ICs have fixed excitatory weights with values such that the output of the ICs is proportional to the mean intensity value over the input. The activation function of the ICs that produces this mean value is a simple weighted sum:

$$v_l = \sum_i C_l(i)\ I(i)$$

where the $C_l(i)$ values are determined by a function that decreases monotonically with distance from the center of the connectable area and sums upto 1. The mean value v_l is used as inhibition to the paired EC, which generates an output according to the equation:

$$u_l = r \cdot \varphi \left[\frac{1 + \sum a_l(i) \cdot I(i)}{1 + \frac{r}{r+1} \cdot b \cdot v_l} - 1 \right]$$

where the weights a_l and b_l are modifiable weights, r represents the efficacy of the inhibitory synapse, and the transfer function is a piecewise linear function according to:

$$\varphi(x) = \begin{cases} x, & x > 0 \\ 0, & x <= 0 \end{cases}$$

The functionality of directional detectors is summarised in Fig. 4.

The directional detectors which have the same directional sensitivity, of neighboring hypercolumns are connected by a link. The directed spreading takes place by these links. Hence the k(R) defined for L1 is sensitive to the direction. The outputs of the layer L1 are connected to the maxima detector. This network is a simple on-center/off-surround network to detect maxima. Each one of the maxima detector cell suppresses the neighboring neurons according to its activation and feeds back excitatory activation to itself.

The second layer L2, also consists of two-dimensional array of neurons. These cells are connected to all their neighbors by links. Each neuron receives its activation from the input and the first layer according to the following equation:

$$L2_{x,y} = I_{x,y} - L1_{x,y},$$

where $L2_{xy}$, I_{xy} is the activation value fed to the neuron of $L2_{xy}$ is the input binary pattern and $L1_{xy}$ is the activation values of L1. From the equation it is clear that the second layer receives complement of the first layer output over the input binary pattern. Since the first layer detects all the lines and diffuses them, the second layer receives activations at corners, curves of all curvatures other than straight lines and contour terminations. In layer L2, the spreading takes place between only the active neighboring neurons. So the corner, curve and contour termination centroids are enhanced. The output of L2 is fed to the maxima detector and the maxima detector locates the enhanced peaks of L2.

Rapid eye-movements(saccades) driven by the bottom-up cues play an important role in the establishment of spatial relations. The absolute and relative positions of the peaks located by L1 and L2 of this system can be considered as bottom-up cues for the eye/camera movement to establish the spatial relationships. The peak strength shows the length oı a line or a curve at that position. The 'on' pixels around the fixed window of the peak is useful for identification of the feature at the peaks. Figure 5a represents the input binary image. Figures 5b-5f represent the outputs of different layers of directed spreading activation layers.

The third layer L3, receives its input from both the maxima detectors of first two layers. This forms the feature map which can be fed for recognition to higher layers. In this layer the conductivity of the region k(R) is constant. Hence the diffusion takes place in all directions. As the diffusion progresses quasi static maxima are formed. These quasi-static maxima identifies the centers of the feature clusters. Eventually global maxima are formed representing the geometric centroid of various feature clusters.

V. CONCLUSION

Spreading activation layer reported in the literature has been used for feature clustering, boundary completion and fixation point generation. In this paper we have shown a new directed spreading activation model. In this model the layers have the capability to detect the low level features like line segments, corners, curves of different curvatures in a purely data driven manner. The feature map generated by this directed spreading activation can be used to direct the eye/camera movement to detect fine features at that location.

REFERENCES:

[BREITMEYER 86] Breitmeyer, B.G. (1986). 'Eye Movements and Visual Pattern Perception,' In E.C.Schwab & H.C. Nusbaum (Eds.) Pattern recognition by humans and machines. (Vol.2, pp.65-86). New York: Academic process.

[COHEN 84]Cohen, M., & Grossberg, S. (1984). 'Neural Dynamics of Brightness Perception: Features, boundaries, diffusion and resonance,' Perceptual Psychophysics, 36, pp. 428-456.

[FUKUSHIMA 82]Fukushima, K. & Miyake, S. (1982). 'Neocognitron: A New Algorithm for Pattern Recognition Tolerant of Deformations and Shifts in Position,' Pattern Recognition, Vol. 15, No. 6, pp. 455-469.

[GROSSBERG 89]Grossberg, S., Mingolla, E., (1989). 'A Neural Network Architecture for Preattentive Vision,' IEEE Transactions on Biomedical Engineering, Vol. 36, No. 1, pp. 65-83.

[HUBEL 77]Hubel, D.H., & Wiesel, T.N. (1977). 'Functional Architecture of Macaque Monkey Visual Cortex,' Proceedings of the Royal Society of London, Series B, 198, pp.1-59.

[RYBAK 91]Ilya A. Rybak, Natalia A.Shevtsova et al., (1991). 'A Visual Cortex Domain Model and Its Use for Visual Information Processing,' Neural Networks, Vol.4, pp.3-13.

[SEIBERT 89]Seibert, M., & Waxman, A.M. (1989). 'Spreading Activation Layers, Visual Saccades and Invariant Representations for Neural Pattern Recognition Systems,' Neural Networks, Vol. 2, pp. 9-27.

[YARBUS 67]Yarbus, A.L. (1967). 'Eye Movements and Vision,' New York: Plenum press.

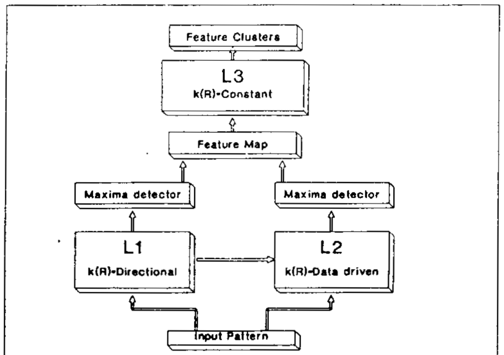

Fig. 1. Directional spreading Neural Architecture

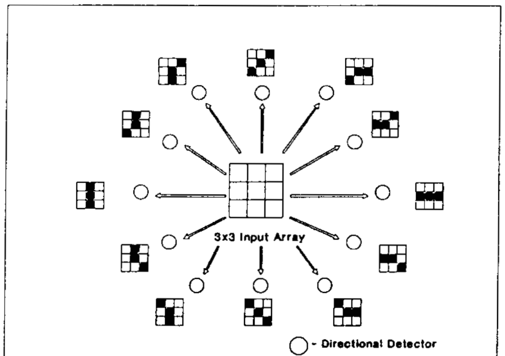

Fig.2 Layer one: Hypercolumn input

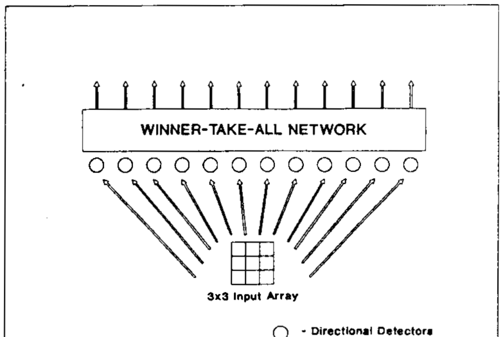

Fig.3 Layer one: Hypercolumn output

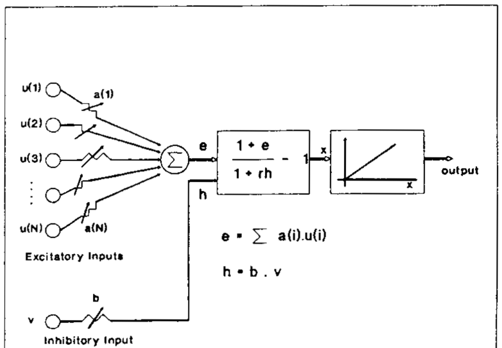

Fig. 4 Characteristics of Directional detector

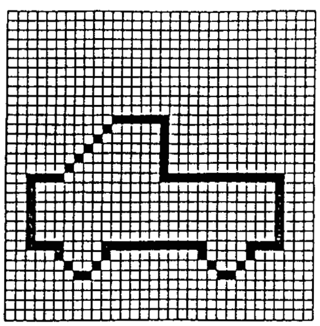

Fig. 5(a) Input Binary Image

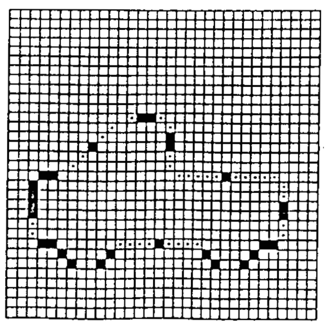

Fig. 5(b) Layer L1 output values

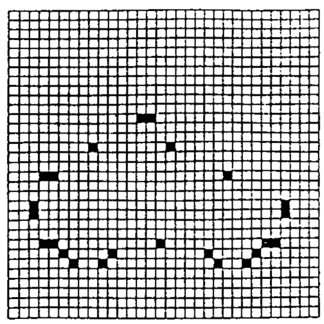

Fig. 5(c) Maxima points in L1

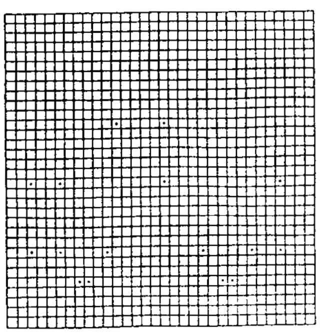

Fig. 5(d) Layer L2 output values

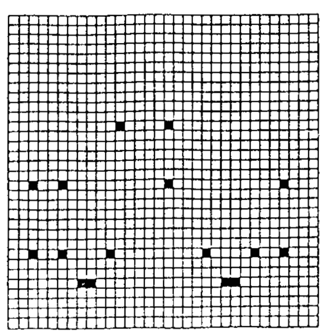

Fig. 5(e) Maxima points in L2

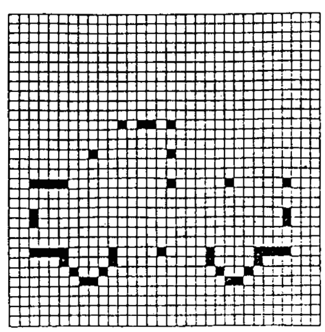

Fig. 5(f) Feature Map for the input pattern