PROCEEDINGS

OF THE

NATIONAL ACADEMY OF SCIENCES, INDIA
1997

VOL. LXVII SECTION–A PART III

Visual Pattern Recognition - Connectionist Perspective

SANKAR K. PAL and JAYANTA BASAK

Machine Intelligence Unit, Indian Statistical Institute, Calcutta-700 035, India.

Received September 23, 1995; Revised August 3, 1996; Accepted February, 13 1997

Abstract

Different classical algorithms for 2D object recognition and the relevance of neural networks are discussed. Connectionist methods are categorized into two groups depending on whether they consider simultaneously the tasks of feature-node association and classification, or not. Research on mixed category perception is emphasized separately in this regard. Some of the limitations of the connectionist approaches are finally listed.

(**Keywords** : Visual pattern recognition/neural works.)

1. Introduction

The term 'visual pattern' refers to a wide variety of cases ranging from the images of microbiological organisms to the satellite images. Visual pattern recognition generally refers to the tasks of finding different meaningful parts (objects) in an image and classifying them according to some predefined object models. In 2D recognition, the planar view of a model is sufficient (e.g., characters, different regions in satellite images, flat objects etc.), while in 3D recognition, a view independent volumetric representation of a model is necessary (which is often difficult to generate).

In this article, we will confine ourselves only with 2D object recognition. For any visual pattern recognition problem there are mainly two stages : *feature extraction* and *feature interpretation*. Feature extraction stage often consists of two major steps : (a) region/edge based feature extraction, (b) secondary feature extraction. This stage follows several preprocessing tasks including enhancement/noise cleaning, segmentation/edge or

line detection and edge/line linking. The features, extracted from an image, constitute a description of the candidate objects to be recognized. In order to identify them the derived features are matched with the selected attributes of stored object models utilizing various techniques. The matching is performed in the *feature interpretation* stage which involves various tasks such as ranking of features, classification, labeling etc.

The feature extraction and interpretation (matching) stages are dependent on each other. Extraction of features is dependent on the type of objects to be found in a scene and consequently the interpretation process to be followed. Similarly, the matching strategy depends on the set of extracted features. For example, if region based features providing some global characteristics of objects are extracted, statistical or distance based decision rules perform better. On the other hand, for structural and relational features characterizing the local properties, AI-based methods[1] are used. The algorithms using the local and relational features have several advantages over those using only global features in many cases like interpreting more than one object simultaneously, dealing with occlusions.

Often twofold processing is performed for interpreting the feature set. One is hypothesis generation which is essentially a bottom-up process to generate a hypothesis about the presence of objects from the extracted feature set. The other is the hypothesis verification, which is a top-down process where the model attributes are matched against those of the image-derived features. Depending on the model descriptions, widely different algorithms are followed for both these processes.

Various classical algorithms based on the theory of statistical/syntactic pattern recognition, AI based techniques (e.g., heuristic search, relational homomorphism, association graph matching, boundary correlation, dynamic programming etc.) or massively parallel computational algorithms (e.g., generalized Hough transform, relaxation labeling etc.) have been developed for object recognition. Fuzzy set theoretic concepts[2,3] have been employed in this regard in order to handle uncertainties at various stages of recognition system arising from vagueness, ill-definedness, incompleteness etc. in input information. Connectionist modeling (based on artificial neural networks, ANNs) which provides an efficient computational framework has also been employed in these tasks extensively. Recently, this paradigm has drawn the attention of researchers from various disciplines.

Another stream of research in connectionist modeling is emerging in the context of simultaneous perception of multiple objects.This may be termed as 'Mixed Category Perception' which essentially refers to the task of simultaneous recognition of multiple entities. The word 'entity' can have different meanings in different contexts like objects in the case of visual pattern recognition, disorders in the case of medical diagnosis etc.

There exist several review articles[1,4-8] concerning the classsical techniques for object recognition, but the connectionist approaches have not been focussed. During the past

fifteen years, a number of attempts has been made towards neural modeling of various aspects of object recognition. This has resulted in various promising methods and methodologies for dealing with these tasks efficiently. Therefore, it seems that a categorical classification of these different connectionist approaches is necessary, at present, for better understanding of the state of the art and furtherance of research in this discipline. In this article, we present different methods and methodologies for object recognition from connectionist point of view including mixed category perception.

2. Classical Approaches and Motivation for using ANNs

Let us now provide a brief outline of the general techniques that are followed in designing the classical algorithms for object recognition. These are association and relational graph matching techniques, heuristic search techniques, boundary correlation based techniques, Hough transform based techniques, and relaxation labeling based techniques.

In association graph based techniques[9,10], a graph is formed by representing the acceptable matches between the scene and model features as the vertices, and connecting the compatible associations by edges (compatibility is determined based on various constraints). After formation of the association graphs, cliques are found in order to obtain the most compatible matches between the objects in the scene and the model objects.

In the techniques based on the relational descriptions of objects[11,12], each object is described by a set of primitives and a set of n-ary relations over the set of primitives. The scene objects (candidates) are matched with the model objects (prototypes) by defining relational homomorphisms, monomorphisms and isomorphisms between the corresponding descriptions. To tolerate noisy and erroneous environment, a concept of $\in$ homomorphism has been fomulated. The problem of finding suitable homomorphism between the shape descriptions are dealt with by general constraint satisfaction tree search.

Hough transform (HT) is often used to extract simple structures[13] like lines, curves of known form, circles etc. The concept of HT was extended to generalized Hough transform (GHT)[14] to match arbitrary contours. In GHT, each edge point in the image is aligned with the model edge points and accordingly, the position of the model object in terms of (x,y,θ) is calculated. Thus for each constituent edge point in each model object, (x,y,θ) values are computed which indicate the plausible locations of the model objects in the image, determined locally by the edge points. A four-dimensional accumulator array $A[N][X][Y][\Theta]$ is maintained where N is the number of model objects, X, Y and Θ are the quantized values of x,y and θ respectively.Corresponding to each (x,y,θ) value computed for each model object, accumulator value is incremented. After considering all edge points in the image, the peaks in accumulator space are found out, which essentially represent the objects' identities and locations. Different variations of GHT are used to recognize the objects from

a scene. Instead of considering all edge points, some characteristic features in the model objects can also be considered.Some algorithms also try to reduce the space requirement to maintain the accumulator array at the cost of inherent parallelism embedded in GHT.

In the heuristic search based techniques, first some estimated positions and orientations of some plausible objects are found from the matched set of features. Then a tree search technique based on heuristic reasoning is employed to find out the match for other features. The nodes in the tree normally represent matches between model primitives and scene primitives. Each node is associated with some weight indicating a measure of similarity of the scene primitives with that of the model. Different variations of A^* search algorithms are usually employed in the heuristic search process.

In the object recognition algorithms based on relaxation labeling technique, the features derived from the scene are initially matched (and then associated) with the model features according to some similarity measures. Then the labelings (associations) are iteratively updated based on some compatibility function which takes care of the labelings of other scene features. The updating process continues till a suboptimal quality of match (may be quantified on the basis of compatibility function) is reached.

Why Neural Networks?

All the techniques, mentioned above, try to formulate the task of object recognition as an optimization (constraint satisfaction) problem. A suitable solution is approached by heuristic search, relaxation lebeling, finding out cliques from association graphs, or finding the homomorphism between relational graphs. The generalized Hough transform is used to obtain some initial guess about the presence of objects. But in order to get the desirable matching performance, the most prominent peak in the Hough space needs to be detected. This, in turn, is a nontrivial task. Again, in the implementation of GHT, space requirement is very high. However, one definite advantage of GHT is that it can be directly implemented on a parallel machine. Also, the degrees of importance of different feature-object pairs can be effectively utilized in GHT.

Whatever methodology be adopted for object recognition, it needs to be robust and fast. Preferably, the algorithms should be implementable on parallel hardware. Second, in these tasks, sometimes it is necessary to associate degrees of importance (or weights) with the features. If the association of importance or weights with the features can be performed adaptively depending on the environment then the methodology may prove to be more versatile.

The performance of the classical algorithms, in general, is not comparable with the real-time performance of the biological systems which are capable of adapting in the environment and seems to be more robust in their behaviour. Principles of animate vision

and their comparisons with machine recognition systems are provided in literature[15-18] Although the objective of machine vision is not necessarily to emulate animate vision, its performance may plausibly be improved if some findings in the fields of neurophysiology and psychology regarding visual cognition can be taken into account in the development of artificial systems. The findings in neurophysiology may provide a bottom-up guideline, while the findings in psychology may provide a top-down guideline for such improvement. Since *artificial neural networks* (ANNs) attempt to provide the closest computational framework of biological nervous systems, the neurobiological and psychological findings may be incorporated into the artificial recognition systems in a better way using ANN models. Moreover, ANNs sometimes provide an alternate framework for dealing with the optimization problems. This does not necessarily mean that artificial neural networks are able to emulate the behaviour of biological systems, rather they sometimes resemble biological systems in a very naive manner. However, the neural networks (or connectionist models) having several basic characteristics like robustness, scope for massive parallelism and capability of learning from examples (adaptivity and generalization capability), provide a tempting paradigm for dealing with the real-life recognition tasks.

3. An Overview of Connectionist Approaches

Connectionist approaches for object recognition sometimes utilize the findings in psychology and neurobiology and try to model the macrolevel properties of the biological systems. Sometimes, the concepts used in these approaches are borrowed from classical techniques for suitable representation and decision making. For example, generalised Hough transform and relaxation labeling have been used for decision making, and association (or relational) graphs have been used for representation in connectionist framework. A suitable amalgamation of these two extremes is also made in certain cases for designing the connectionist systems.

The existing approaches broadly consider two different aspects of object recognition. One is the proper representation of features and objects in the connectionist framework, and the other is decision making. Depending on whether the networks consider these aspects together or not, they are divided in two categories (Fig. 1). Category 1 deals only with the decision making part where the task of mapping of features is not embedded into connectionist framework, rather it is performed separately. The connectionist decision making is performed based on the principles of classification or clustering or optimization, and depending on the type of decision making, different network models are employed. The networks in Category 2, on the other hand, consider both mapping of features and decision making part within connectionist framework. The additional mapping circuitry for feature-node association plays the role of human intervention (as needed in Category 1). Note that, the

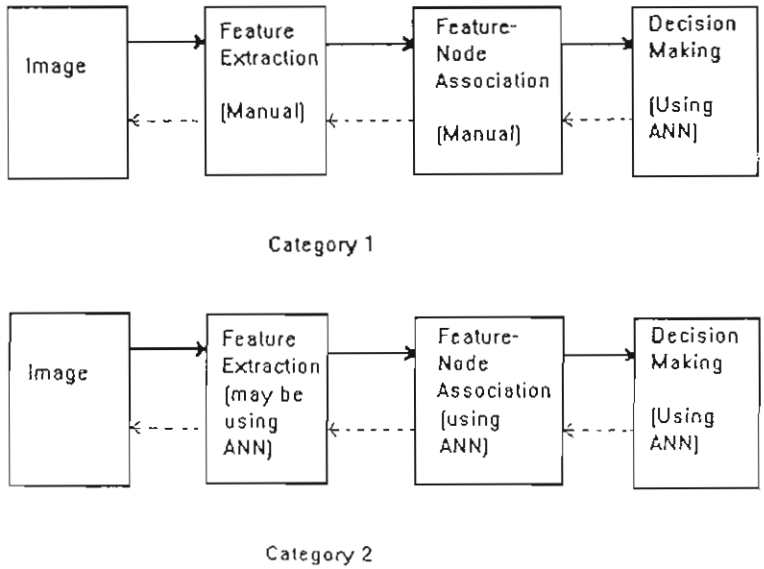

Fig 1 – Different stages involved in categories 1 and 2. The dotted lines indicate the possible existence of feedback pathways.

design criteria of the mapping circuitry is dependent on the specific application at hand. For this reason, we refer to the models under this category as application-specific systems.

In category 1 (Fig. 1), generally some basic neural networks like MLP, Hopfield model, Kohonen model, and ART are used for classification/ clustering the input feature vectors. The extracted features from a scene are suitably mapped onto the input of the networks by the user. In category 2, application- specific systems are built based on some indigenous and/or basic networks in order to adopt different techniques for automatically mapping the features onto the networks as well as decision making. These mapping techniques are sometimes motivated from neurobiology, sometimes from psychology and many times they are analogous to the classical concepts. Let us now briefly describe them.

3.1 *Connectionist Decision Making (Category 1)*

Here we describe a few methods along with their principles using Hopfield model,MLP, and self-organizing network.

3.1.1 Hopfield Model Based Methods

Principle: Hopfield model has been used in many object recognition algorithms by posing the task as an optimization problem (the capability of Hopfield model to deal with

optimization problems was first shown by Hopfield and Tank[19]). Here, an energy function is defined for the network such that the minima of the function corresponds to optimal match between the image features and the model features. The dynamics of the network is set in such a way that the network settles to a local minima of its energy function through iterative updating of the states of its units, and the state vector corresponding to the local minima provides a suboptimal match between the image and model features. Despite the fact that the Hopfield net is not guaranteed to provide the optimal solutions, rapid computational capability of the network provides an important mechanism for tackling the computational complexity involved in dealing with object recognition problems. However, a limitation of using Hopfield model is that the degrees of importance of different feature-object pairs cannot be automatically associated, i.e., the weights cannot be learned. Rather, these weights are preassigned before any matching task is performed.

Techniques : Nasrabadi and Li[20,21] developed a scheme for object recognition with the help of a Hopfield model where the polygonal approximations of 2D objects (both the model and scene objects) were represented as graphs. A Hopfield network with two-dimensional array of neurons (number of rows represents the number of scene features and number of columns represents the number of model features) was used to match the model graphs (one at a time) with the scene graph. The best matching subgraphs correspond to the objects present in the scene. However, this particular method dealt with only symmetric relations between features, i.e., matching between only undirected graphs was performed. Features using 'sphericity' property of objects[22] were also used to form the graphs in this scheme.

A Hopfield net based technique was developed for matching structural descriptions of objects (descriptions of parts and spatial relations between them)[23]. Here, a transformation of the shape descriptions was used such that shape descriptions containing asymmetric spatial constraints between the parts can be matched using symmetric interconnection weights for the Hopfield net.

The task of 3D object recognition using Hopfield net was performed by Lin *et al.*[24] Here, the network was used for matching prototype objects with the stored model objects in two stages : feature-wise in the first stage and surface-wise in the second stage. Here also, only the symmetric relations between features were considered.

3.1.2 Multilayer Perceptron Based Methods

Principle : Multilayered perceptron has been used for object recognition by posing the task as a classification (supervised) problem. Here, feature vectors are derived for each object model and an MLP is trained with these feature vectors under supervised mode. The trained network then accepts the features extracted from the image and classify them accordingly.

Techniques : In the method proposed by Tsang *et al.*[25] for object recognition using MLP, the features were extracted as follows. The entire range of angles (0° - 360°) was divided into a number of slots of equal size, and each slot has been treated as a feature. The feature value was determined by the number of corners whose angles fall within the corresponding slot. Besides this, another feature vector was formed in a similar way, where a feature value was equal to the number of arc changes (between two successive corners) by an angle within the corresponding slot. The union of these two feature vectors then applied as input to the network for learning and classification.

The system developed by Tsang and Yuen[26] for the recognition of partially occluded objects, had three stages. In the first stage, after detection of object boundaries, some salient features were extracted from a feature-codebook. In the second stage, presence of some possible objects were hypothesized using a nonlinear elastic matching technique. Finally, the presence of each possible object was verified using the corresponding salient features with the help of an MLP. However, in this system, MLP was used only in the final stage, and hypotheses about the presence of the objects were made in the nonlinear elastic matching process itself.

Bebis and Papadourakis[27] devloped an MLP based recognition system where some invariant features were extracted by using the cumulative angular and curvature representations of the object boundaries.

3.1.3 Self-Organizing Net Based Methods

The principle used in this methods are analogous to the MLP based methods, except that the decisions are made in unsupervised manner (i.e., without the help of any external teacher). Bebis and Papadourakis[27] also investigated the effectiveness of Kononen model with the same features for performing the task of object recognition.

Adaptive resonance theory has also been used for shift and orientation invariant visual pattern recognition. Srinivasa *et al.*[28] used an invariance network in conjunction with an ART1 module for this purpose. Here, different rotations (0°, 90°, 180°, 270°) and shifts (in discrete steps) were explicitly coded in the invariance network which cooperatively interacts with the output layer of ART1 module. Although shift and rotation invariance were achieved for simple form of visual patterns, real-life objects were not considered as input. Moreover, ART1 module also has the same diffculty, like MLP, in the perception of mixed categories.

3.2 *Application-Specific Systems (Category 2)*

In order to explain the principles utilized in Category 2, let us consider here five different application-specific systems namely, Neocognitron[29,30], DHT model (*dynamic Hough transform*)[31,32], MORSEL (*multiple object recognition and selective attention*)[33],

TRAFFIC (*transforming feature instances*)[34] and PsyCOP (psychologically motivated connectionist system for object perception)[35]. These systems generally accept the input image directly or sometimes the spatially distributed features (e.g., edge map, skelatal version of the image etc.). The general difficulty of designing such systems lies in the incorporation of the characteristics : translation, rotation, and scale invariance into the systems. Again the incorporation of these invariance properties depends on the way of mapping the features automatically onto the network. This is also dependent on the type of objects to be recognized by these systems. Here we specify three different strategies, generally followed for this purpose, which are

1. employing layers of analyzers (a collection of processing elements/neurons) to achieve invariance,

2. computing rigid transformations (i.e., shift, orientation, and scaling transformations) explicitly within neural architecture,

3. using selective attention mechanism.

In strategy 1, the analyzers hierarchically extract and group the features from an image. The lower layer analyzers extract simpler features while the higher layer analyzers extract more complex features by grouping the simpler features extracted in the lower layer. The analyzers in each layer, are able to respond to the patterns with a small amount of invariance, because the neurons in each layer accept activations from a group of neurons in the lower layer. Thus, in this technique, the invariance is achieved incrementally within the layers of analyzers. This particular strategy is motivated by the theory of perceptual organization and also by the organization of optical nervous system of the animals. This strategy also enables a system to tolerate a good amount of deformations besides shift, orientation, and scale change. This kind of strategy is useful for the recognition of alphabets, numerals etc. and their deformable versions. The application-specific systems like Neocognitron, MORSEL, employ such strategy.

In strategy 2, the transformations from the feature reference frame to the object reference frame are computed within the connectionist systems. In this process, the principle of GHT is often incorporated within the links of a connectionist architecture so that for a given input object, the maximum activation goes to a particular neuron representing the identity, position, orientation, and scale of the object. Sometimes, the activation values of the neurons are updated in a way similar to the relaxation algorithms. The incorporation of rigid transformations is performed by using different connectionist learning algorithms including backpropagation. Sometimes, learning rules are used to generate the internal representations alongwith their reference frames in the neural architecture. This kind of strategy for computing transformations is particularly useful in the recognition of rigid objects (e.g., industrial objects) and is exploited in the systems like TRAFFIC, DHT model and PsyCOP.

The word 'selective attention' itself explains the fact that a zone in the image space is selectively attended at a time. This psychological phenomenon is sometimes incorporated into neural architectures (strategy 3) in order to map certain portions of the image selectively onto the input layer of a recognition system. An additional mapping circuitry (attention controller) is often employed for this purpose. Whenever a particular zone is attended, the attention controller has the relevant information about the location of that zone. Thus, if some object is present in the zone of attention, the positional information in the attention controller can aid the recognition system to specify the position of this object. The selective attention mechanism (apparently a sequential process) is helpful in building connectionist systems which are used for reading texts/scripts or even for recognizing overlapped objects. This particular mechanism is employed in many of the application-specific systems including MORSEL, PsyCOP, and enhanced versions of Neocognitron.

3.3 *Comparison of Characteristics*

In object recognition techniques based on the basic networks like Hopfield, MLP, or Kohonen model (category 1), the descriptions of the objects needs to be explicitly derived by some other algorithms, and these descriptions are then fed to a neural achitecture. In other words, the neural architectures, in this kind of applications, behave as coprocessors in the main recognition systems. In these techniques, there should always be some expert intervention or some separate process which properly maps the features or relational descriptions onto a network model. The application-specific systems, on the other hand, take either the segmented image or spatially distributed features and use them directly as input. It is true that in some of these systems, features need to be extracted separately, but mapping of the features onto the network is performed by the systems themselves and no separate expert intervention is therefore necessary. In other words, application-specific networks are aimed more towards building up the stand-alone systems for object recognition.

Hopfield model has been used for multiple object recognition by matching model graphs one at a time with the scene graph, but the problem of mixed category perception (i.e., simultaneous recognition of multiple objects) has not been dealt with. Moreover, the degree of importance of the features cannot be learned.On the other hand, feature importance can be learned with MLP or Kohonen model based techniques. Here also, the problem of mixed category perception cannot be handled. Hopfield model based techniques are similar in nature to the relaxation labeling techniques used in the classical algorithms whereas the techniques based on MLP are similar to the statistical or decision theoretic rules in pattern recognition.

The application-specific systems mainly differ from the points of representation of features and capability of decision making. The architecture of Neocognitron is motivated by the hierarchical structure of visual cortex in order to perform hierarchical grouping of

Table 1 – Comparison of the characteristics of application-specific systems.

Characteristics	Neocognitron	DHT Model	TRAFFIC	MORSEL	PsyCOP
Object type	numerals, characters	alphabets	astral constellation	words	Industrial objects
Feature type	pixel	strokes, junctions	geometric features	line segments	corners
Translation invariance	yes	yes	yes	yes	yes
Rotation invariance	no	yes	yes	no	yes
Scale invariance	yes	no	yes	no	no
Multiple objects	no	?	yes	yes	yes
Multiple instances (same object)	no	?	no	yes	yes
Control	BU/TD	BU/TD	BU	BU/TD	BU/TD
Rigid transformation	no	yes	yes	no	yes
Learning	supervised (*)	—	supervised	supervised	supervised
Selective attention	yes	no	no	yes	yes
Psychological evidence	yes	no	no	yes	yes
Psychological explanation	?	yes	no	yes	?

features (strategy 1). TRAFFIC employs strategy 2 (which is analogous to the classical concept of generalized Hough transform) to explicitly compute the position, orientation, and scaling information of an object. DHT model also employs this strategy, but the location, orientation, and scaling information are not explicitly computed, rather the spatial locations of the activated neurons provide this information. MORSEL, on the other hand, employs selective attention mechanism (strategy 3) for locating different letter clusters. PsyCOP integrates both the selective attention mechanism and the generalized Hough transform technique in its architecture. These five application-specific systems incorporate the invariance properties to different extents. For example, Neocognitron and MORSEL do not take care of the orientation invariance. Similarly, PsyCOP, in its present form, is not able to perform scale invariant recognition.

As far as the capability of decision making is concerned, the task of simultaneous recognition of multiple objects has not been considered in most of the systems (except MORSEL and PsyCOP). Although TRAFFIC is able to take care of multiple objects in some limited sense, it is not able to recognize multiple instances of the same object. MORSEL not only incorporates some of the psychological findings (e.g., selective attention) but also interprets a few disorders of psychological patients. PsyCOP has been essentially developed based on the psychological finding that identification and localization occur in two separate zones of the visual cortex.

The characteristics of five different systems are summarised in Table 1. In this table, *a* (?) mark indicates questionable performance. For example, in the task of multiple object recognition, DHT model faces problems of finding one object in some other's location if the parameters of the network are not properly tuned. 'BU' indicates a bottom-up control while 'TD' indicates a top-down control. Neocognitron was originally designed only with a bottom-up control, but later top-down control has also been incorporated in order to identify the most prominent one from a mixed set of visual patterns. TRAFFIC uses only bottom-up process (i.e., the verification from object layer to the feature layer is not performed), while DHT model, MORSEL, and PsyCOP use both bottom-up and top-down controls. Neocognitron is generally trained only with supervised mode, but there is a theoretical formulation for unsupervised learning also. *A* (?) mark in the psychological explanation indicates that it is not clear whether any psychological phenomenon can be explained or not with these models.

The connectionist models described so far, particularly those under the framework of 'category 1', are able to recognize one object at a time. In real-life, we often encounter with situations where more than one object appear simultaneously and sometimes they can also occlude each other. In such cases, it is necessary to recognize multiple objects simultaneously.It is very difficult to deal with the task of simultaneous recognition of multiple objects under the framework of 'category 1' using MLP, Hopfield model, Kohonen's

model, or ART model. For example, if we use MLP then the feature set extracted from an image consisting of more than one object (even if they do not occlude each other) is essentially an overlap of the feature sets corresponding to different objects. As a result, the input feature vector falls widely apart from the true decision regions formed by the parameters learned with the examples from individual classes. Due to the similar reason, it is also very difficult to directly use self-organizing models for such tasks. In Hopfield net based methods, the task of recognition of single object is modeled as a problem of finding the most suitable match for a candidate feature set between a number of prototype feature sets. In the case of multiple objects, another degree of complexity is added to this task of optimization. The problem becomes even more complicated in the case of occlusion and missing features.

In order to deal with the task of simultaneous recognition of multiple objects, several other connectionist models[36-51] have been developed. Although many of these models have been developed for different purposes other than object recognition (e.g., predicting multiple disorders of a patient from his symptoms, perceiving music from multiple sources etc.), their principles can be properly exploited in the task of object recognition. In the next section, we discuss the principles of these models under the framework called, 'mixed category perception'. We then explain how the principle of mixed category can be applied to real-life object recognition under both categories 1 and 2.

4. Mixed Category Perception : Principle and Models

Ideally (under noiseless condition), the problem of mixed category perception can be described as follows. Let m different objects (O), characterized by the collections of different features (f), be represented as

$$O_1 = \{f_{11}, \ldots, f_{1n}\}$$

$$O_m = \{f_{m1} \ldots\ldots, f_{mn}\}$$

Let a new feature vector $F_k = \{f_1, \ldots, f_N\}$ be formed by the superposition of k different objects, i.e.,

$$F_k = \bigcup_{i \in \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, m\}} O_i$$

Then the task of mixed category perception is to identify a set of k objects $\{O_{j1}, O_{j2}, \ldots, O_{jk}\}$ such that

$$F_k = \bigcup_{j \in \{j_1, \ldots, j_k\} \subseteq \{1, \ldots, m\}} O_j = F_k .$$

Note that, it is desirable to find out exactly the same set of k objects which were superposed to form the feature vector F_k. However, this is possible if F_k is resulted from a unique combination of k objects. Otherwise, another set of k objects may be found out which, when superposed, would result in the same F_k. Thus, mixed category perception appears to be equivalent to 'set covering' problem.

4.1 *Models for Mixed Category Perception*

In literature, there exist several investigations for mixed category recognition from various points of view including the models developed by Peng and Reggia, Cho and Reggia, masking field, EXIN, SONNET and X-tron. In the model developed by Peng and Reggia[52], possible disorders are predicted for a given set of manifestations. However, it has no learning scheme, and the connection weights are fixed depending on the conditional probability values of the diseases given the symptoms are present. Cho and Reggia[36,37] developed a supervised learning scheme (error back-propagation) for automatically assigning these weights through competition and cooperation processes. In EXIN (an acronym for excitatory and inhibitory connections)[44-46] the inhibitory strengths between the competing nodes are updated in such a way that the competition process gets confined within the output nodes of similar nature, i.e., getting activations from inputs which have sufficient overlap. This is performed by strengthening the connection weights between coactivated neurons and weakening the connections to inactive neurons.Thus the network achieves a limited form of mixed category perception. In masking field[38,39] (self-similar, gain-controlled, cooperative-competitive feedback network), the coding property is adaptively sharpened with the repetitive presentations of a pattern at the input. In SONNET (an acronym for self-organizing neural network)[40-43] output nodes responding to similar input patterns compete between themselves, which is essentially a similar concept used in EXIN. However, SONNET has a novel property of forming stable codes for embedded patterns.

Recently, a three-layered connectionist model, called X-tron[47-51] (Fig. 2) was proposed for simultaneous recognition of multiple categories/objects. Let us now describe its principle and key features in brief. The input layer accepts numerical values representing the degree of presence of features. The output layer produces the degree of presence of categories/objects, and the hidden layer corresponds to the feature-object associations.

Whenever an input pattern in presented to the network, the activations are propagated to the hidden layer and then through the bottom-up links the activations reach the output layer. The output layer, in turn, feeds back the activations to the hidden layer. Each hidden node is activated for a particular input-output combination. The hidden nodes connected to a common input node compete with each other.The winner-take-all node physically represents the strongest possible hypothesis (i.e., the most probable object) associated with

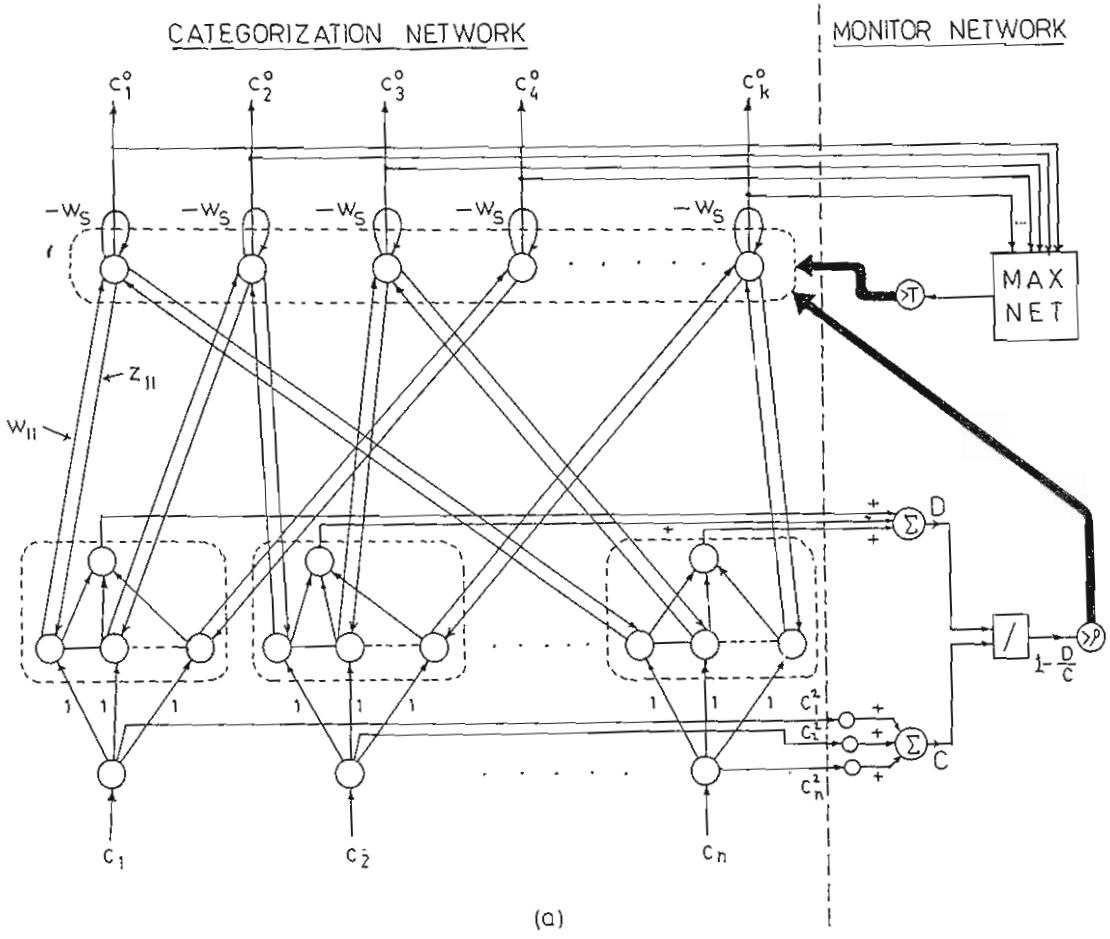

Fig. 2 – Structure of the connectionist model for self-organization. Bold lines represent the control paths from the monitor network to the categorization nettwork. Another node is added in each box of hidden nodes which measures the ambiguity at the corresponding feature (considering the set of hidden nodes). Two adders are used in conjunction with hidden and input layers to compute D and C respectively. The node denoted as "/" computes the certainty factor $1 - D/C$.

that particular input feature. The difference in activations of the input node and the WTA hidden node is again propagated through the bottom-up link connected to the WTA node. Each output node has a negative self-feedback. The feedback ensures that in the process of updating, the activation of an output node automatically decreases if it does not get support from its constituent features. If an output node gets proper support from its constituent features (i.e., the botom-up differential support and negative self-feedback cancel each other) then its activation will stabilize to some nonzero value. These outputs actually yield a measure indicating how well the feature set is being interpreted by the network. If the activation to a particular input node does not match with the support received from the output categories, then there is an ambiguity corresponding to that feature. If a single pattern or a set of mixed patterns from a new class appears, the features are not fully interpreted by the learned categories. If the total ambiguity for all the features is greater than some threshold then the pattern presented to the network is considered to belong to a new category. With each hidden node there is another node which computes the ambiguity at that particular node. These ambiguities and the input activations are propagated to the monitor where the certainty factor is calculated. The certainty factor is considered to have a linear relation with the total ambiguity. If the certainty factor is greater than some threshold then the pattern is considered to be already present, and the weights of the links are iterated. If the certainty factor is less than the threshold then a new output mode is allocated for the input pattern. This threshold is called the *vigilance threshold* (ρ) which is supplied externally. This vigilance threshold is analogous to the vigilance factor used in adaptive resonance theory[53].

In the monitor network there is another part which checks the maximum activation present at the output layer. The maximum activation is checked if it is greater than a threshold T. If the condition holds true then only the network is allowed to learn the pattern present at the input. Otherwise, even if the certainty factor is greater than the vigilance threshold the learning is not allowed. Whenever a new node is allocated in the output layer the upper part of the monitor network is activated and the network learns the new pattern.

In X-tron, instead of using selective form of competition (as performed in EXIN or SONNET), the competition process is degenerated from the output layer. The learning rules for X-tron are defined in such a way that the weights of the links asymptotically reach some previously defined probabilistic measures. During learning, the network automatically adjusts the number of nodes in the hidden layer. X-tron is able to learn both under supervised and unsupervised modes.

The learning rules or the rules for iterative adjustment of connection weights are set in such a way that the weight of each link asymptotically reach a previously defined measure. The measure of each weight is defined in such a way that it achieves the ability to capture

the relative frequency of appearances of the corresponding feature object pairs. The asymptotic measures are

$$z_{li} = p\,(f_i/o_l) \tag{1}$$

and

$$w_{il} = \frac{p\,(o_l/f_i)\,p\,(f_i/o_l)}{\gamma + \sum_{i=1}^{n} p\,(o_l/f_i)\,p\,(f_i/o_l)} \tag{2}$$

Here w_{il} represents the weight of the bottom-up link from the $(i,l)^{th}$ hidden node (connecting i^{th} input node and l^{th} output node) and the l^{th} output node. Similarly, z_{li} is the weight of the corresponding top-down link. The constant γ is used to get the effect of Weber's law.

The learning rules are

$$\frac{dw_{il}}{dt} = \left(\alpha_i\,\delta_l\,z_{li} + \alpha_l^o\left(\frac{w_{il}}{z_{li}}\right)\right) c_i\,y_l - (\alpha_i\,c_i + \alpha_l^o\,y_l)\,w_{il} \tag{3}$$

and

$$\frac{dz_{li}}{dt} = \alpha_l^o\,y_l\,(c_i - z_{li}) \tag{4}$$

where α_i and α_l^o are the agility factors of the i^{th} input node and the l^{th} output node. The agility factor determines the capability of learning of the links connected to that node. Higher the agility factor, higher will be the rate of learning and *vice versa*. Initially, the agility factors of all nodes are set to unity and they are decreased with the learning trials. The agility factor of a hidden node is the same as that of the input node connected to it. The agility factors are changed according to the following rules.

$$\frac{d\alpha_i}{dt} = -\,\alpha_i^2\,c_i\,. \tag{5}$$

$$\frac{d\alpha_l^o}{dt} = -\,\alpha_l^{o2}\,y_l \tag{6}$$

The value of y_l is the desired output value of l^{th} output node. The desired output is determined by the monitor network (if supervised mode is used then it is supplied externally). The value of δ_l is given as

$$\delta_l = \frac{\varepsilon_l}{\gamma g'(u_l)} \tag{7}$$

where $\varepsilon_l = y_l - o_l$ is the difference of the actual output at the l^{th} output node from its desired value (as determined by monitor network). γ is a constant used in the asymptotic measure (equation 2) which also controls the rate of learning (as expressed in the learning rules). The LTM equations for top-down links are similar to those presented by Grossberg[53,54].

4.2 *Principle of Mixed Category Perception Applied to Object Recognition.*

In Section 3, we have seen that the models like MLP, Kohonen's model, Hopfield model, or ART model can be used both in Category 1 and Category 2 for object recognition. In a similar way, the models for mixed category perception can be applied as basic building blocks in the development of connectionist systems for simlutaneous recognition of overlapping objects in the framework of both Categories 1 and 2 (Fig. 1). As an example, let us consider PsyCOP, mentioned in Section 3.2, in order to illustrate how the principle of X-tron has been employed there for simultaneous recognition of overlapping 2D objects under Category 2.

The architecture of PsyCOP (Fig.3) consists of two parts, one for decision making and the other for associating the image features with the input nodes of the decision making network. The decision making network (consisting of three layers) is designed based on the architecture and principle of operation of X-tron. In X-tron, input nodes represent only the numerical values of the features, but in order to associate the image features with the input nodes, as required in PsyCOP, it is essential to have the positional information of the features also. For this purpose, PsyCOP employs two different channels: one to represent the features/objects (i.e., 'what it is' part) and the other to keep information regarding the location of features/objects (i.e., 'where it is' part). The neurons in these two channels are connected through interchannel links. Note that, the existence of two separate channels is also found in the animal brains which is supported by the psychological evidences[55,56].

Like X-tron, the competitive process in the hidden layer decides the categories which finally get activated in the output later. Note that, the links (Fig.3) from input to hidden layer store the information about the transformations from feature reference frame to object reference frame.

The mapping circuitry employs selective attention mechanism with the help of an attention control network which scans the image sequentially, and then through the control

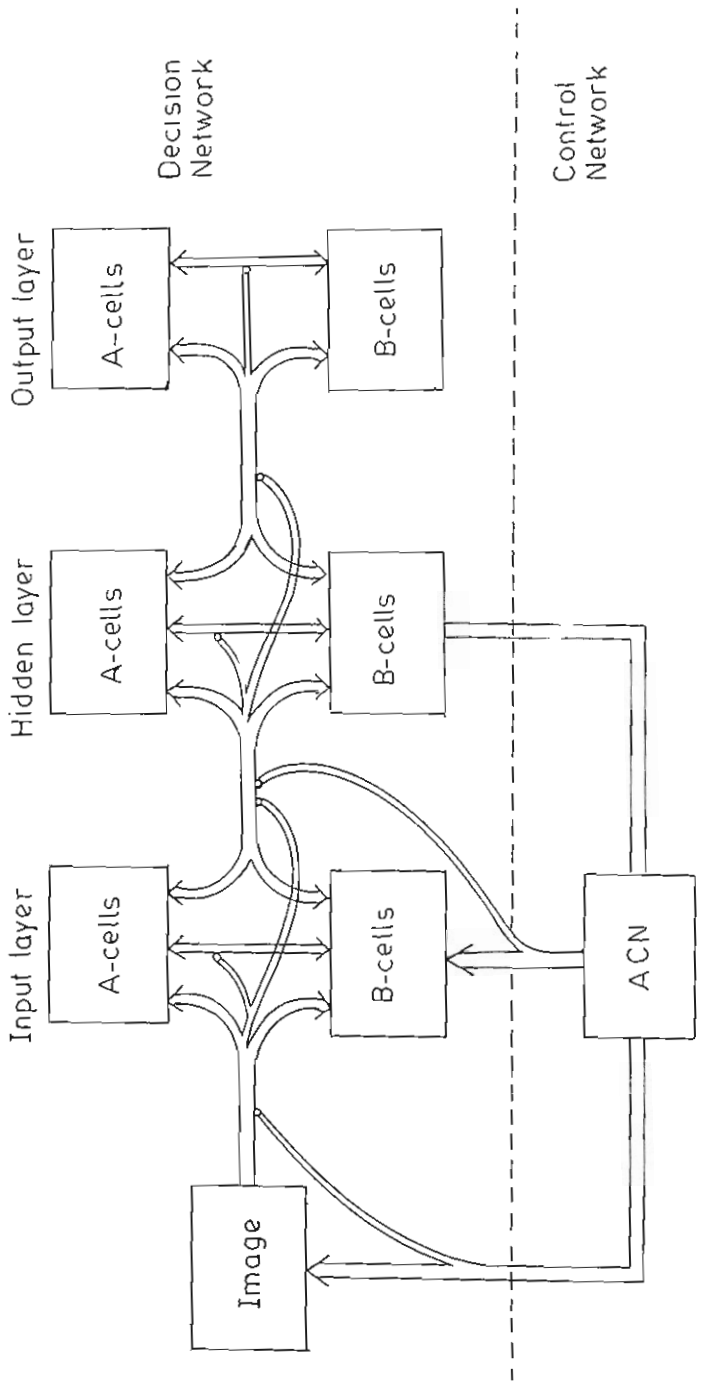

Fig. 3 – Block diagram of PsyCOP. The decision network determines the location and identity of an object while the control network controls the selective attention mechanism

network the corner features trigger the input neurons of the identification and localization channels. Different input neurons in the identification channel are tuned to different cornerity values (here, angles of the corners are used) and the input neurons in the localization channel are tuned to different possible locations of the features (i.e., these neurons cumulatively cover the entire image space). The activation patterns of the input neurons in the identification and localization channels are coordinated by the attention control circuit. The neurons in these two channels form a stable coalition to represent where a particular feature or category is located.

5. Conclusions and Discussion

Different connectionist methods for object recognition, in a unified framework, is presented alongwith a survey on mixed category perception. The different classical algorithms for 2D object recognition are also briefly discussed in this respect. Connectionist approaches are classified based on the principles that some of the approaches consider only the decision making part (category 1), while the others consider both decision making and automatic mapping of the features within connectionist framework (category 2).

Although neural networks have several merits, they find some limitations and/or problems in dealing with the task of real-life object recognition. Some of them are mentioned below.

- *Architecture - selection of the proper architecture for decision making :*
  Although different architectures have been proposed, it is difficult, even for a specific problem, to decide which particular configuration of a model would perform better. It is also sometimes difficult to choose an appropriate model for a given application domain.

- *Mapping - proper mapping of the features and subparts onto the network with their spatial distribution*:
  It is always necessary to properly transform a real-life task in terms of the variables acceptable by the computational model of a neural network. Sometimes it may be a pretty hard problem.

- *Binding - proper representation of the identity and location together and proper representation of the context:*
  For object recognition, it is necessary to represent the information on 'what' and 'where' of a subpart or a feature (or sometimes the entire object) together, but efficient design of such representation may be difficult. Moreover, in some cases the interpretation of the subparts (features) may change depending on the context. For example, a particular shape/object can have different meanings in different

context. Representation of this kind of knowledge is still found to be difficult in the connectionist framework.

- *Hardware : design of the hardware for real-time performance:*
  Massive parallelism, a characteristic feature of ANN, would become apparent only when these architectures would have suitable hardware realization or at least can be efficiently simulated on a parallel machine. The main bottleneck of NNs hardware design is the implementation of variable synaptic connections. During the learning phase, NN are supposed to change their weights frequently which is difficult to realize in the available hardware systems.

Furthermore, it is not always possible and appropriate to compare the performances of connectionist and classical approaches. This is because many NN based approaches are developed not exactly to provide better results for specific tasks, but to explore their generic merits (e.g., parallelism, robustness, learning ability etc.) and to give better insights to Psychological and neurological findings. For example, with the help of MORSEL, an empirical study has been made to explain neglect dyslexia of psychological patients[57]. psyCOP provides a phenomenological model of visual cortex based on the behavioral characteristics. Neocognitron tries to explore the layered nature of visual cortex. Again, connectionst approach on neural modeling often provides an alternate paradigm for looking at a problem from different points of view. However, sometimes it is found that the classical AI based approaches provide more appealing results for certain specific tasks[4,5,7].This is due to the fact that such AI algorithms are mostly goal-driven, i.e., they are developed for very particular types of objects with specific domain knowledge.

In conclusion, a hybrid system can thus be designed within a structured connectionist framework (with a richer knowledge representation scheme which may be borrowed from classical concepts) incorporating some of the psychological and neurological findings for better architectural efficacy. The system should be able to learn and recognize multiple objects simultaneously.

Acknowledgment

This work is supported by CSIR, New Delhi under project grant No. 22(235)/93/EMR-II.

References

1 Wallace, A. (1988) *Pattern Recognition* **21** : 241.

2 Pal, S K. & Majumdar, D Dutta (1986) *Fuzzy Mathematical Approach to Pattern Recognition*, Wiley, New York

3 Bezdek, J.C. & Pal, S.K. (1992) *Fuzzy Models for Pattern Recognition: Methods that Search for Structures in Data* IEEE Press, N.Y.

4. Suetens, P., Fua, P. & Hanson, A. (1992) *ACM Computing Surveys* **24** : 5
5. Besl, P. & Jain, R. (1985) *ACM Computing Surveys* **17** : 75.
6. Chin, R.T. & Dyer, C.R. (1986) *ACM Computing Surveys* **18** : 69
7. Zhao, F. (1991) *Int. J. Pattn. Recogn. Artificial Intell.* **5** : 715.
8. Chaudhury, S. (1989) *Ph. D. thesis*, Deptt. of Computer Science and Engg., Indian Institute of Technology, Kharagpur, India
9. Bolles, R.C. & Cain, R.A. (1983) *Robot Vision*, ed., A. Pugh, Springer, Berlin
10. Kashyap, R. & Koch, M. (1985) *Proc. IEEE Conf. on Robotics and Automation, St. Louis*, p. 150.
11. Shapiro, L. & Haralick, R. (1982) *IEEE Trans. Pattern Anal. Mach. Intell.* **4** : 595.
12. Shapiro, L. & Haralick, R. (1985) *IEEE Trans. Pattern Anal. Mach. Intell.* **7** : 90.
13. Illingworth, J. & Kittler, J. (1988) *Computer Vision, Graphics and Image Processing* **44** : 87.
14. Ballard, D.H. (1981) *Pattern Recognition* **13** : 111
15. Bullock, B.L. (1978) *Computer Vision Systems*, eds. Hanson, A.R. and Riseman, E.M., Academic Press, New York.
16. Hochberg, J. (1987) *Computer Vision, Graphics and Image Process.* **37** : 221.
17. Skryzpek, J. (1989) *"Neural specification of a general purpose vision system,"* Tech. Rep. CSD/890072, Computer Science Department, University of California, Los Angeles, USA.
18. Ballard, D.H. & Brown, C.M. (1992) *CVGIP : Image Understanding* **56** : 3
19. Hopfield, J.J. & Tank, D.W. (1985) *Biological Chybernatics* **52** : 141.
20. Nasrabadi, N. & Li, W. (1991) *IEEE Trans. Syst., Man and Cybern.* **21** : 1523.
21. Li, W. & Nasrabadi, M. (1989) *IEEE Int. J. Conf. Neural Networks, II*, Washington, DC, p. 287.
22. Ansari, N. & Li, K. (1993) *Pattern Recognition* **26** : 531
23. Basak, J., Chaudhury, S., Pal, S.K. & Majumdar, D. Dutta (1993) *Int. J. Patt. Recogn. Artificial Intell.* **7** : 377.
24. Lin, W., Liao, F. & Lingutla, T. (1991) *IEEE Trans. on Neural Networks* **2** : 84
25. Tsang, P.W.M., Yuen, P.C. & Lam, F.K. (1992) *Pattern Recognition* **25** : 1167.
26. Tsang, P.W.M. & Yuen, P.C. (1993) *IEEE Trans. Systems, Man and Cybernetics* **23** : 228.
27. Bebis, G.N. & Papadourakis, G.M. (1992) *Pattern Recognition* **25** : 25
28. Srinivasa, N. & Jouaneh, M. (1993) *IEEE Trans. Systems, Manand Cybernetics* **23** : 1432.
29. Fukushima, K. & Miyake, S. (1982) *Pattern Recognition* **15** : 445
30. Fukushima, K. (1988) *IEEE Computer*, p. 65
31. Hilton, G. (1981a) *Proc. IJCAI-81, International Joint Conf. for Artificial Intelligence*, Vol. 2.
32. Hilton, G. (1981b) *Proc. IJCAI-81, International Joint Conf. for Artificial Intelligence*, Vol. 2.
33. Mozer, M. (1991) *The perception of multiple objects: A Connectionist approach*, MA: MIT Press, Cambridge.
34. Zemel, R. (1989) *Tech. Rep.* CRG/TR/89/2, Dept. of Computer Science, University of Toronto, Canada.
35. Basak, J. & Pal, S.K. (1995) *IEEE Trans. Neural Networks* **6** : 1337.
36. Cho, S. & Reggia, J. (1993) *Neural Computation* **5** : 242
37. Reggia, J., DAutrechy, C., Sutton III G. & Weinrich, M. (1992) *Neural Computation* **4** : 287.

38. Cohen M & Grossberg, S. (1986) *Human Neurobiology* **5** : 1.
39 Cohen, M. & Grossberg, S. (1987) *Applied Optics* **26** : 1866.
40. Nigrin, A. (1990) in *Proc. Int. Joint. Conf. Neural Networks*, Washington DC, p. 525.
41. Nigrin, A. (1990) in *Int Joint.Conf Neural Networks*, San Diego, CA , p 313.
42. Nigrin, A. (1990) *Ph D. Thesis*, Duke University.
43 Nigrin, A (1992) in *Proc. Int. Joint. Conf Neural Networks*, Baltimore,MD, p. 683
44 Marshall, J. (1990) in *Proc. Int. Joint. Conf Neural Networks*, San Diego, CA, p 649.
45 Marshall, J (1990) in *Proc Int. Neural Networks Conf.*, Paris, France, p. 809.
46. Marshall, J (1992) in *Proc. Int Joint. Conf. Neural Networks*, Baltimore, MD, p 315
47 Basak, J., Murthy, C.A , Chaudhury, S & Majumdar, D. Dutta (1992) in *Proc. Eleventh IAPR Int. Conf on Pattern Recognition*, Hague, Netherlands, p. 36
48. Basak, J , Murthy, C.A , Chaudhury, S & Majumdar, D Dutta (1993) *IEEE Trans. Neural Networks* **4** : 257.
49. Basak, J., Murthy C A & Pal, S.K. (1996) *Neurocomputing*, **10** 341.
50 Basak, J. & Pal, S K (1993) in *Proc INSA-CSI Seminar on Pattern Recognition, Artificial Intelligence and Neural Networks*, Dehradun, India, p 5
51 Basak, J & Pal, S.K (1995) *IEEE Trans Neural networks* **6** . 1091.
52. Peng, Y. & Reggia, J (1989) *IEEE Trans Systems, Man and Cybern.* **19** : 285.
53 Carpenter, G.A. & Grossberg, S (1987) *Computer Vision, Graphics and Image Processing* **37** : 34.
54. Grossberg, S (1982) *Studies of Mind and Brain*, Reidel Press, Boston
55. Kosslyn, S M (1975) *Cognitive Psychology*, **7** . 341.
56. Kosslyn, S., Holtzman, J , Farah, M & Gazzaniga, M (1985) *Journal of Experimental Psychology: General*, **114** : 311.
57 Mozer, M. & Behrmann, M. (1989) *Tech Rep* CU-CS-441-89, Dept. of Computer Science, University of Colorado, Boulder, USA.