

PK M. S. I. 18

OBJECT RECOGNITION FROM DEGRADED IMAGES USING NEURAL NETWORKS

A *THESIS*

submitted for the award of the degree

of

MASTER OF SCIENCE

in

COMPUTER SCIENCE AND ENGINEERING

by

A. RAVICHANDRAN



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

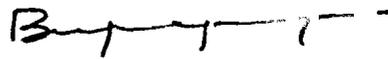
INDIAN INSTITUTE OF TECHNOLOGY

MADRAS-600 036, INDIA

MARCH 1993

CERTIFICATE

This is to certify that the thesis entitled **OBJECT RECOGNITION FROM DEGRADED IMAGES USING NEURAL NETWORKS** is the Bonafide work of **Mr.A.Ravichandran**, carried out under my guidance and supervision, in the Department of Computer Science and Engineering, Indian Institute of Technology, Madras, for the award of the degree of **MASTER OF SCIENCE** in Computer Science and Engineering.



(B.Yegnanarayana)

ACKNOWLEDGEMENTS

I thank my Guide **Prof.Yegnanarayana** for initiating me into the discipline of research and patiently guiding me through the course of my thesis with his amazing acumen. It is my childlike desire to emulate his dedication to research, his mental alertness and flexibility of his thinking besides his technical intuition. He has provided us with an excellent environment to grow.

I thank the **Department of Electronics, Govt. of India**, for supporting me in the initial stages of this thesis through the sponsored research project, "Algorithms for Signal Reconstruction in Acoustic Imaging".

Ramaseshan, besides introducing the fundamentals of computer science, has served as a model student for me. I thank him for his contributions towards this thesis through countless discussions. I thank **Dr.Chouhan** for the open-mindedness and enthusiasm **with** which he has helped me shape my research effort. **Sundar, Hemn, Chandra and Ramana** have helped me with their thorough reasoning abilities and receptiveness whenever I approached them with doubts.

Arul has shared many deep and disturbing questions, about things technical and nontechnical, along with corresponding un-answers. Together, we have **unlearnt** many things. I also thank him for **organising** my otherwise sphagetti C-code.

Uday, Sai and Murali through countless fiery discussions on everything under earth (including the present research) have contributed to many ideas contained herein. Uday has taught me the value of poise, patience and planning.

Rajendran, Madhukumar and **Ramachandran** have benefitted me through illuminating technical discussions in topics other than my thesis. I have no words to thank my roommate Prabha for his constant support and his suggestions to improve my thinking and communication skills. I thank **Alwar** for several interesting discussions.

I thank my parents, Meenakshi and Annaswamy, for giving me all freedom during the course of my education and encouraging me to explore and learn.

CONTENTS

ABSTRACT

CHAPTER 1

INTRODUCTION	1
1.1 Object recognition from degraded images	2
1.1.1 Sensor array imaging context	2
1.1.2 Need for an object recognition system	4
1.2 Need for adaptive processing and Neural networks	5
1.2.1 Shortcomings of traditional pattern recognition techniques	5
1.2.2 Need for adaptive processing	6
1.2.3 Artificial neural networks	6
1.3 Overview of the thesis	8
1.3.1 Motivation	8
1.3.2 Scope	8
1.3.3 Thesis organisation	9

CHAPTER 2

REVIEW OF APPROACHES TO TRANSFORMATION INVARIANT OBJECT RECOGNITION	11
2.1 Review of approaches to transformation invariant object recognition	12
2.1.1 Traditional approaches	12
2.1.2 Neural approaches	14
2.1.3 Recent approaches : the hybrids	16
2.1.4 Summary of findings	17
2.2 Limitations of approaches in the context of degraded images	17
2.3 Plan of the present research effort	18
2.3.1 Overall approach	18
2.3.2 Aspects of complexity	19
2.3.3 Choice of symbol set	19

2.3.4 Study in three stages	19
-----------------------------	----

CHAPTER 3

OBJECT RECOGNITION FROM DEGRADED IMAGES 20

3.1 Difficulties in Object recognition from degraded images	20
3.2 Correlation matching using a neural network	21
3.3 Hamming network	21
3.3.1 Structure	21
3.3.2 Learning	21
3.3.3 Classification	23
3.4 Experimental studies	23
3.4.1 Effect of increasing sparsity	25
3.4.2 Performance with data collection at multiple frequencies	25
3.4.3 Effect of noise	25
3.5 Results and Discussion	34
3.6 Summary	35

CHAPTER 4

TRANSFORMATION-INVARIANT RECOGNITION OF OBJECTS 36

4.1 The problem of transformational variability	36
4.2 Approaches to solve the spatial transformations problem	37
4.2.1 Single stage neural network approaches	37
4.2.2 Invariant feature approach	37
4.3 The invariant feature space	38
4.3.1 Theory of moments	39
4.3.2 Algebraic moment-invariants	40
4.4 The classifier	41
4.4.1 Classification	41
4.4.2 Multilayer neural networks	43
4.4.3 Error back propagation algorithm	44

4.5 Experimental studies	46
4.5.1 Experiments in Classification	46
4.5.2 Experiments on the feature space	50
4.5.3 Experiments on the neural network classifier	50
4.6 Discussion	51
4.7 Summary	52
CHAPTER 5	
TRANSFORMATION INVARIANT RECOGNITION OF OBJECTS FROM DEGRADED IMAGES	54
5.1 Introduction : Need for preprocessing	54
5.2 Preprocessing for Noise suppression and Object Extraction :	
Our approach	56
5.3 A Neural Network for Preprocessing	57
5.4 Experimental studies in Preprocessing and Classification	60
5.4.1 Data set	60
5.4.2 Preprocessing performance	61
5.4.3 Classification performance	62
5.5 Discussion	66
5.6 Summary	68
CHAPTER 6	
SUMMARY	69
6.1 Summary of the thesis	69
6.2 Major results of the thesis	70
REFERENCES	72
LIST OF FIGURES	75
LIST OF PUBLICATIONS	79

ABSTRACT

In many practical situations it is required to recognise an object from poorly resolved and noisy images irrespective of changes in scale, position and orientation. Example situations include sensor array imaging, locating buildings in aerial photographs and industrial inspection. Variability of objects in the image pattern and image degradation make model based pattern description and matching difficult.

The objective of this thesis is to explore the possibility of developing *artificial neural network models* that can be trained to identify objects from poorly resolved, noisy and transformed, (scaled, rotated and translated) images, such as images reconstructed from sparse and noisy data. Our aim is to exploit the adaptive processing capabilities of neural networks, both as classifiers and as context sensitive processors.

Studies reported here were made by simulating sparsity of data and noise as obtained in a simplified model of a *sensor array imaging* situation. Sparsity of data is due to limited number of sensors and noise is added to this data due to medium disturbances and unwanted sources. Noise and sparsity of data in the imaging context result in degradation of the quality of the reconstructed image as a whole instead of affecting it in the form of local corruption of the image pixel information as in many image processing contexts. Hence (i) **neighbourhood** processing methods for noise cleaning may not be applicable, (ii) feature extraction cannot be reliably performed and (iii) model based methods for classification cannot easily be applied. We show in this thesis that neural network models can overcome some of these limitations by their learning and context sensitive processing capabilities.

We describe studies in object recognition for three different cases:

In the first case we consider the issues of noise and sparsity alone. Here we describe *object recognition from degraded images* using a simple trained neural network like Hamming network. We show that if the shapes of expected objects are known, then it is possible to train a neural network for object recognition. Even though in this case the network performs a simple correlation matching, it classifies even images

which are so degraded that we fail to perceive discriminating features visually. However, this approach is useful only in situations where the image is not transformed.

In the second case we consider the issue of *transformations* alone. We describe a feature space that is invariant to transformations and a neural network classifier for recognition. We show that a two stage approach, where the tasks of transformational invariance and learned classification are handled separately, is successful in recognising objects over a wide range of scales and orientations. We also show that if an invariant feature space is available, a multilayer neural network classifier can learn object shapes without explicit description.

In the third case we address the issues of noise and transformations together and study *transformation invariant recognition of objects in the presence of noise and sparsity*. We show that a neural network based *preprocessing stage* can be used to overcome the effects of noise and sparsity. Transformation invariant object recognition can then be performed from the processed images. The proposed neural network architecture for preprocessing uses context sensitive lateral cooperation and competition between nodes with receptive fields of various sizes to achieve noise suppression.

In this study, we have considered the following factors responsible for degradation of images of objects in the sensor array imaging context: sparsity of data, noise in the received data, transformational variability and increased detail in the image. We have examined situations with increasing complexity along each of these factors and when they occur together. Results show that unlike model based techniques which have to be tailored to specific types of degradation, neural networks perform well under different types of degradation. More importantly, with increasing complexity due to increased detail in the image **and/or** due to degradations in imaging, neural network based methods appear to exhibit a *gradual degradation in performance*. In situations where complexity is less, these systems can be used as standalone object recognition systems. As the complexity of the recognition task increases, the output of the system can still be used as an aid to human decision making.

1

INTRODUCTION

In many practical situations it is required to **recognise** an object from degraded images. Recognition of an object may also be required irrespective of changes in scale, position and orientation of the object in the image. Example situations include sensor array imaging, locating buildings in aerial photographs, and industrial inspection. In this thesis, we address the problem of recognition of objects from degraded images obtained through reconstruction from sparse and noisy data, as in the case of sensor array imaging. In Section 1.1, we introduce the context of sensor array imaging and stress the need for an object recognition system.

Conventional pattern recognition techniques, which use model based pattern description and matching, cannot be applied here due to the presence of various **types** of degradation. This necessitates adaptive processing schemes that employ learning and context sensitive processing. *Artificial neural networks* have been shown to exhibit these adaptation characteristics. They also exhibit parallel and distributed processing and ability to perform collective computation. In Section 1.2 we discuss the need for adaptive processing techniques and introduce artificial neural networks

The *objective of this thesis* is to explore the possibility of using artificial neural network models that can be trained to identify objects from degraded images, if the set of the expected object shapes is known.

1.1 OBJECT RECOGNITION FROM DEGRADED IMAGES

Degradation of an image may occur during its acquisition due to several reasons. In this section we describe how image degradation occurs in the sensor array imaging context. We stress the need for an object recognition system.

1.1.1 Sensor Array Imaging context

The aim of a sensor array imaging [YEGNANARAYANA 91] system, such as an underwater acoustic imaging system, is to obtain an image of an object by transmitting energy and sensing the wave radiation reflected from the object using a sparse array of sensors as shown in Fig. 1.1.

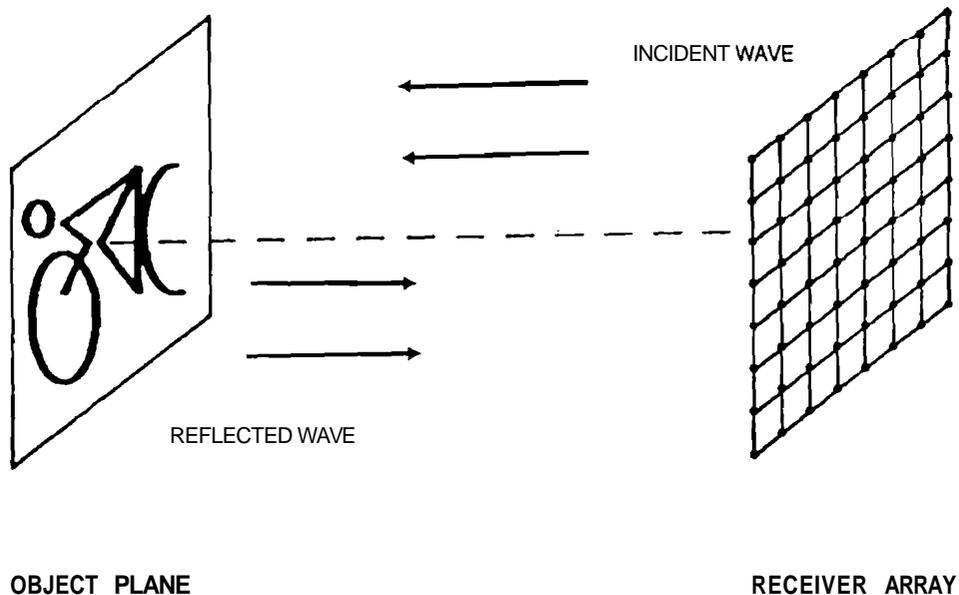


Fig. 1.1 A simplified Sensor Array Imaging setup.

It should be noted that such an imaging system differs from conventional photography in three important aspects:

- (i) It does not employ a lens-like mechanism. As a result, image formation is holographic, *i.e.*, each point on the object plane contributes to every point in the image plane, and the sensed data is a transformation of the image of the object. To obtain an image, the process of image formation is modelled on a computer and the image is computed from the data collected using the equations for image formation. This process is known as *image reconstruction*.
- (ii) The counterpart of a photographic film is also absent. Instead we have an array of sensors that collect reflected energy at the sampling points. Theoretically, images of arbitrarily high resolution can be obtained by employing a receiver array of sufficiently high receiver density. But in practice building and operating dense arrays is very difficult. Usually a *sparse* receiver array containing very small (compared to the resolution of the object in the image) number of sensors is used. As a result, the reconstructed images are *poorly resolved*.
- (iii) *Noise* in the medium may be added to the received data causing further degradation of the reconstructed image. This noise, because it is added to the receiver data which is in a transform domain, has its degrading effect distributed all over the image. This situation is different from many image processing contexts where noise may cause local corruption of image pixel information.

Thus due to sparsity of data and noise, the reconstructed images are poorly resolved and noisy. It is **difficult** to **recognise** an object from reconstructed images due to lack of visual clues required for recognition. The task of recognition becomes much more complex when the object moves relative to the receiver array, in which case, recognition is **required** independent of changes in scale, position and orientation of the object in the image.

We **illustrate** these issues with an example. We have chosen a set of 20 olympic games symbols as planar objects for recognition studies (the reasons for the choice of this set **will** be discussed in a later section). **Fig.1.2(a)** shows some olympic game

symbols (128x128 points) used in simulation studies. Images(128x128 points) of these objects reconstructed from the data collected by an array consisting of 16x16 sensors are shown in Fig.1.2(b). This illustrates image degradation due to data sparsity and noise in the imaging context.

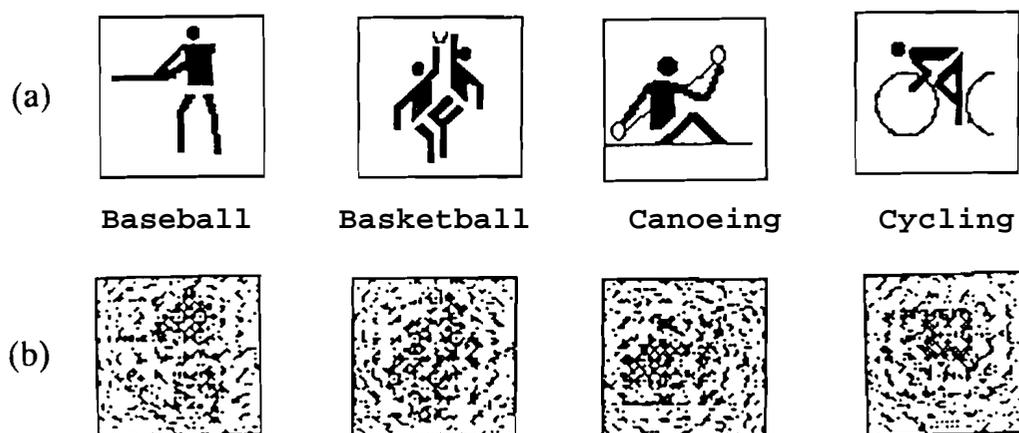


Fig.1.2 (a) Some olympic game symbols (128x128 points) used as planar objects in Sensor Array Imaging simulation studies (b) Corresponding images (128x128 points) reconstructed from data collected by a sparse (16x16 sensors) array.

1.1.2 Need for an object recognition system

As the number of expected targets becomes larger and larger, it becomes difficult for a human observer to identify the object from such poor quality images. Hence there is a need to develop a knowledge based recognition system. It is reasonable to expect that in a noisy environment the performance of a machine may be superior to that of human beings because:

- (i) The machine may extract and use information which may be quite different in nature from what a human expert uses.
- (ii) Human performance degrades with increasing number of targets because one has to remember all of them and apply them consciously, whereas the **performance** of a machine does not depend on the size of the set of

symbols (128x128 points) used in simulation studies. Images(128x128 points) of these objects reconstructed from the data collected by an array consisting of 16x16 sensors are shown in Fig.1.2(b). This illustrates image degradation due to data sparsity and noise in the imaging context.

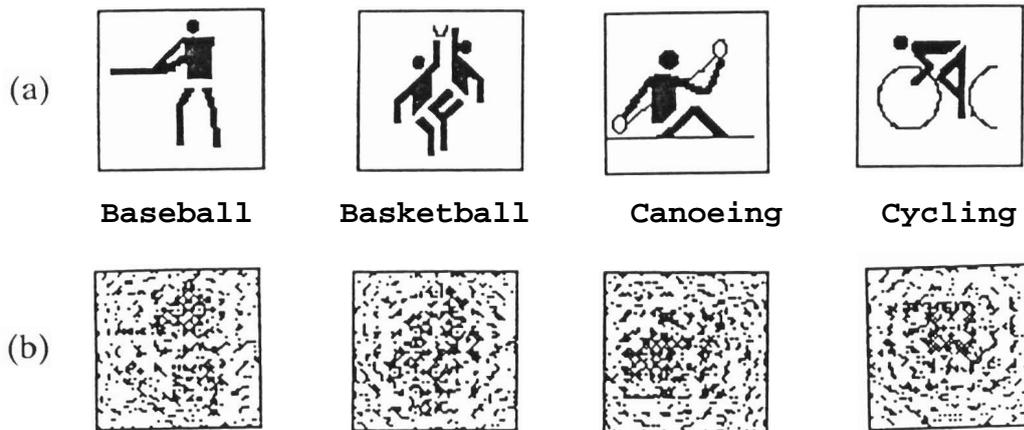


Fig.1.2 (a) Some olympic game symbds (128x128 points) used as planar objects in Sensor Array Imaging simulation studies (b) Corresponding images (128x128 points) reconstructed from data collected by a sparse (16x16 sensors) array.

1.1.2 Need for an object recognition system

As the number of expected targets becomes larger and larger, it becomes difficult for a human observer to identify the object from such poor quality images. Hence there is a need to develop a knowledge based recognition system. It is reasonable to expect that in a noisy environment the performance of a machine may be superior to that of human beings because:

- (i) The machine may extract and use information which may be quite different in nature from what a human expert uses.
- (ii) Human performance degrades with increasing number of targets because one has to remember all of them and apply them consciously, whereas the performance of a machine does not depend on the size of the set of

the targets. Similarly it does not degrade due to fatigue caused by prolonged effort.

- (iii) Contrary to common belief, human beings are very poor in recognising an image if it is rotated considerably. For example, while a simple geometric shape is correctly identified even when rotated, a familiar human face is almost impossible to **recognise** if it is turned upside down.

Thus a knowledge based system is desirable for reliable, quick and accurate recognition of objects from noisy and partial input images.

1.2 NEED FOR ADAPTIVE PROCESSING AND NEURAL NETWORKS

In this section we briefly outline traditional pattern recognition techniques and their shortcomings when data is uncertain. We discuss the need for adaptive processing and introduce **artificial** neural networks.

1.2.1 Shortcomings of traditional pattern recognition techniques

Traditional methods of pattern recognition can be broadly classified into two categories: template matching techniques and structural description techniques.

Template matching techniques, historically the earliest, are used for recognition of printed numerals and characters of the alphabet. Here **all** the exemplar patterns representing the various classes are stored as templates and the pattern to be **recognised** is matched pixel by pixel with every template. The class label corresponding to the template with maximum correlation is chosen as the recognised class.

In complex problems like **multifont** character recognition and recognition of simple geometrical objects, it may not be possible to list beforehand all possible **exemplars**. Matching is costly in terms of computation time and space. The *syntactic or structural description approach* tries to overcome this limitation. This regards images as composed of a fixed set of constituent entities or primitives according to certain rules. Thus, possible compositions are defined implicitly by the rules of composition. Given

the pattern to be recognised, we identify the constituent atoms and then describe how they are put together. This description is then matched with that of all the class examples and the best suited description is chosen as the class label.

In the presence of pattern variability and noise even the syntactic approach may fail. For example, in the case of handwritten character recognition, it is difficult to identify or specify a finite and fixed set of primitives. Moreover, the transformation each primitive undergoes and the various ways in which these transformed primitives are placed in relation to each other may vary widely. There is no closed form solution to the problem of **identifying** the constituent atoms and arriving at a description. Hence the logical structure of a description is too difficult to aspire in noisy and uncertain situations.

To summarise, traditional techniques for pattern recognition *model* pattern variability and design suitable representations and techniques. In practical situations, pattern variability may be too complex to be explained using simple models and it is seldom due to any one type of degradation but rather is the compound effect of many factors simultaneously present. Hence both the data and the models are uncertain.

1.2.2 Need for adaptive processing

In uncertain situations, it is desirable that the processing mechanism adapts its structure and functions. Adaptation has two aspects:

- (i) *Adaptive acquisition of knowledge about the task domain* : Representations and descriptions suitable to the task domain must be **acquired** in the long term by **learning from** examples.
- (ii) *Adaptive application of the knowledge* : Data must be processed by activating the knowledge in a **context sensitive** manner. This is a short term process.

1.2.3 Artificial Neural Networks

The human brain employs both these **types** of adaptation and appears to effortlessly perform the tasks of pattern recognition which remain unsolved even

partially by a machine. In an effort to understand how this is possible, studies about the structure and function of the animal brain have been made [BALLARD 82][FISCHLER 87]. The power and versatility of the brain seems to arise from the fact that the brain is a huge network of highly interconnected neurons, each of which is a processing unit. The following features have been highlighted to account for the performance of the brain:

A. Structural Features:

- (i) Parallel and distributed processing, which gives rise to the advantages of speed, fault tolerance and noise immunity.
- (ii) Analog and nonlinear computing structures, which enable nonsymbolic processing and the use of multivalued logic.

B. Functional Features:

- (i) Modularity and hierarchy, exhibiting clear cut division of labour, which resolves issues of control and coordination among parallel and distributed processors.
- (ii) Learning capacity, which is the ability to build suitable internal knowledge representations from examples.
- (iii) **Self organisation** of structure and function tuned to the realworld problem domain.

Computational models inspired by these studies, which attempt to emulate certain simplified features of the brain are called artificial neural networks. Artificial neural networks have been shown to exhibit some of the desirable characteristics of the brain at least to a limited extent in simple problems [GROSSBERG 80][LINSKER 88][LIPPMANN 87] [RUMELHART 86A]. Research literature abounds with applications of artificial neural networks which exploit their learning and **optimisation** capabilities. In this thesis, we shall describe how simple networks which exhibit the above said features can be put to use in a complex pattern task.

1.3 OVERVIEW OF THE THESIS

1.3.1 Motivation

In this thesis, we study systems which use neural network principles in performing the difficult task of **recognising** objects from degraded images obtained by reconstruction from sparse and noisy data. The aim is more towards evolving techniques which address recognition problem in all its complexity and fail gracefully in performance as complexity increases, rather than build a system whose performance is tuned to a simplified context and may fail with complexity. Thus in situations where the complexity is less, the system can work as a stand-alone object recognition system. As complexity of the problem increases due to increased detail in the **image** and/or due to imaging degradation, the output of the system can still be used as an aid to human decision making.

1.3.2 Scope

Studies reported here are made by simulating noise and sparsity as obtained in a simplified model of a sensor array imaging situation. We limit the study to simple metric **transformations**(**rigid** object transformations in 2-dimensions), namely, scaling, translation and in-plane rotation. Occlusion and **perspective** projection issues have not been addressed since they involve models of three dimensional shape of the object.

The following issues are addressed in this thesis: (i) We have studied systems for **object recognition from poorly resolved and noisy images**, such as images reconstructed **from** sparse and noisy data. (ii) We have studied the problem of **transformation invariant recognition of objects from noisy images** for varying levels of complexity. We have shown that in each case it is possible to build recognition systems that perform successfully **upto** a level of degradation. As the problem becomes complex the performance degrades gradually. (iii) We have demonstrated the **effectiveness of neural networks** both as classifiers and as context sensitive adaptive processors, by addressing issues of data uncertainty and variability due to noise and distortion.

1.3.3 Thesis organisation

In **Chapter 2** we review earlier approaches for transformation invariant recognition of objects, both nonneural and neural. We discuss their applicability and limitations in the present context. Finally we outline the research plan. In the following chapters we describe our studies in object recognition for three different cases as shown in Fig.1.3.

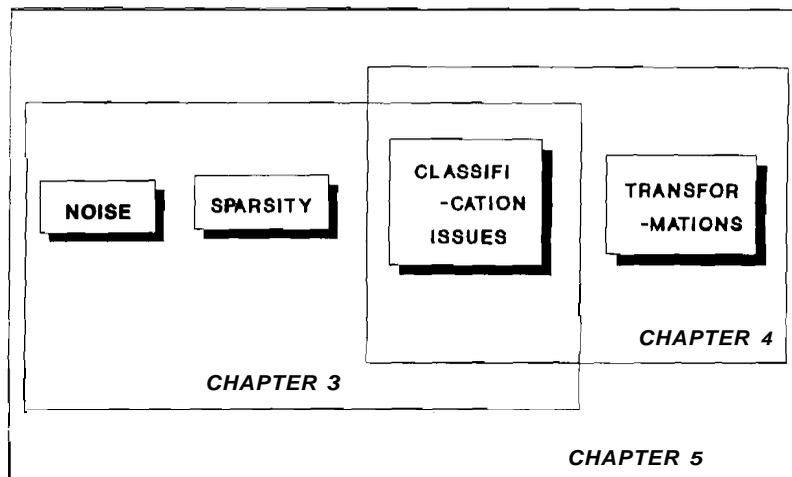


Fig.1.3 Overview of the research showing simplified situations where selected issues are considered.

In the first case discussed in **Chapter 3**, we consider the issues of noise and sparsity alone. Here we describe object recognition from degraded images using a simple trained neural network.

In the second case, we consider the issue of transformations alone. We describe this in **Chapter 4**, where a feature space that is invariant to transformations is designed and a neural network classifier is used for recognition.

In the third case discussed in **Chapter 5**, we address the issues together and study transformation invariant recognition of objects in the presence of noise and sparsity. In

the first two cases existing neural network architectures are explored for recognition of objects. In the third case some new neural network based methods are proposed for preprocessing.

In Chapter 6 the results are summarised.

2

REVIEW OF APPROACHES TO TRANSFORMATION INVARIANT OBJECT RECOGNITION

Transformation invariant recognition of complex man made objects like submarines, **aircrafts** and industrial products began to gain attention in 1970's. Almost **all** early systems used a suitably designed invariant feature space along with a traditional classifier. Neural approaches for transformation invariant recognition were actively pursued in 1980's. In these approaches, attempts were made to obtain **invariance** either by suitably designing the structure or by training the networks. Since 1990 hybrid systems which use invariant feature spaces to handle transformation and neural networks for classification are being studied. We review these techniques in brief in Section 2.1.

However, transformation invariant object recognition from degraded images is a much more complex task. In the case of images reconstructed from sparse data, the task of recognition becomes complex due to various types and varying levels of degradation. In Section 2.2 we discuss the limitations of earlier approaches in the present context.

In Section 2.3 we outline our research plan.

2.1 REVIEW OF APPROACHES TO TRANSFORMATION INVARIANT RECOGNITION

2.1.1 Traditional approaches

Though transformation invariant recognition of patterns related to complex man-made objects began to gain attention only in 1970's with increase in applications in industrial inspection, identification of military vehicles, aircrafts and submarines, the first significant work considering moments for invariant recognition of objects was performed by Hu as early as 1962.

A.Moment Invariants using Cartesian moments

Hu [HU 62] derived combinations of moment values that are invariant with respect to scale, position and orientation based on theories of invariant algebra. He also demonstrated the utility of moment invariants using a simple pattern recognition experiment using a set of 26 capital letters as input patterns. He used the first two moment invariants to represent the digitised patterns of these letters in a two dimensional feature space. An unknown pattern could be classified by computing its first two moment values and finding the minimum Euclidean distance between the unknown and the set of known patterns in feature space. When plotted in two dimensional space, **all** points representing each of the characters were distinct. However, some characters that were very different in image shape were very close in feature space. In addition, slight variations in input images of the same character resulted in variations in feature values that lead to overlapping of closely spaced classes. **Hu** concluded that increased image resolution and a larger feature space would improve object distinction.

Dudani et al [DUDANI 77] applied moment invariants to a model based three dimensional object recognition system, whose goal was to perform automatic classification of **aircrafts from** television images. Moment invariant feature vectors were computed **from** silhouette and boundary information. It was claimed that high frequency details in the image are best characterised by moments derived from the

object boundary while overall shape characteristics are best represented by silhouette moments. Object classification was based on a distance weighted k-nearest-neighbour rule between the object feature vector and the feature vectors of the model database. Their results showed the moment based classifier to be more accurate than several qualified human observers.

Alternative moment invariant techniques as well as variations of Hu's proposal have been suggested. Maitra [MAITRA 79] presented a variation of Hu's moments that is additionally invariant to contrast change and inherently size invariant.

B. Rotational moments

Rotational moments based on a polar coordinate representation of the image were proposed as they have well defined rotation transform properties. However, they have complicated translation transformations. Hence rotational moment techniques rely on Cartesian moments to find the center of mass and then compute the rotational moments about that point.

Smith and Wright [SMITH 71] used a simplified rotational moment technique to derive invariant features for characterising images of ships. Yin and Mack [YIN 81] compared the effectiveness of rotational invariants with Cartesian (Hu's) invariants for object classification and found that both provided similar results. However, it was observed that Hu's features **require** less computation time than rotational moments.

C. Orthogonal Moments

Orthogonal moments, which are obtained by replacing the nonorthogonal monomial basis set with an orthogonal basis set (e.g., **Legendre** and **Zernike** polynomials) were proposed by Teague [TEAGUE 80]. **Legendre** moments are orthogonal moments defined over Cartesian coordinates and **Zernike** over polar coordinates.

The advantages of these orthogonal moments is that a small set of these can be used to reconstruct an image to an approximation. The following are the drawbacks of orthogonal moments:

- (i) A large number of **higher** order moments may be required to sufficiently characterise an image for a given application.
- (ii) A large number of features may not carry any useful information.
- (iii) Computationally they are much more expensive than Hu's invariants.
- (iv) Translation invariance of the Zernike moments is poor, being defined over polar coordinates.

Khotanzad and Hong [KHOTANZAD 90] used a set of rotationally invariant features based on the magnitudes of **Zernike** moments in recognising 24 English characters. 23 **Zernike** features, when used with a nearest neighbour classifier, performed better than moment invariants.

Teh and Chin [TEH 88] performed an extensive analysis and comparison of the most common definitions of moments. In terms of sensitivity to additive random noise, high order moments are the most sensitive to noise. Some studies have implied that important information may be contained in the high order moments, whereas in most practical experiments, there is little improvement in identification performance when moment orders are increased beyond order 4 or 5. An excellent review and survey of moment based techniques is provided by [PROKOP 92]

2.1.2 Neural approaches

Biological systems are extremely adept at some forms of invariant recognition. Several researchers have attempted application of artificial neural networks to invariant pattern recognition, with the hope that the characteristics vital to invariant classification can be abstracted and utilised in simplified networks. In this section we review some of these approaches and discuss their limitations.

There exist at least two classes of techniques for invariant recognition of objects using an all-neural architecture. First, the *structure* of the network can be designed such that its output is always invariant to certain transformations. Alternatively, representatives of a large class of transformations can be presented to the network during *training* so that it learns which transformations are equivalent.

A. Invariance by Structure

Obtaining invariance in the response of a network by appropriately designing its structure has been proposed by several authors [FUKUSHIMA 83] [GILES 88] [WIDROW 88] [WAIBEL 89].

In all these approaches, one creates connections between neurons which force transformed versions of the same input to have the same output. For example, let us consider an input image which is to be classified independently of in-plane rotations about its center. Let w_{ij} be the weight leading to a neuron n_j from the pixel i in the input image. Rotational invariance can be achieved by imposing the following structural constraint: $w_{ji} = w_{jk}$ for all i and k which lie at equal distances from the center of the image. In this case, rotation of the input image will not change the total input to any neuron.

This example can be used to point out the limitations of these approaches. The same weight w_{ji} has to be duplicated for every pixel at the same radial distance from the origin. Therefore the number of connections required for images of realistic size is extremely large. For example, it has been estimated for one such classifier [WIDROW 88] that for an input image consisting of N pixels, to provide complete transformational invariance, N^4 connections are necessary. In practice, images have to contain at least 10^4 pixels for satisfactory resolution and hence this approach is not realistic from the engineering point of view.

B. Invariance by training

Here, the classification ability of neural networks is used directly to obtain transformation invariance. To do this, a number of different examples corresponding to different transformations of the same object are input to the neural network for training. If the network is able to learn to discriminate these objects properly, and if the number of examples shown is large enough, the network may generalise correctly to transformations other than those shown. This approach was taken by Rumelhart et al [RUMELHART 86B] to obtain invariance for the 3x3 objects which they investigated.

From an engineering perspective, invariance by training has two disadvantages:

- (i) The notion that a network trained to recognise one object invariantly can use this training to recognise new objects invariantly is not clearly understood or proved. With present techniques, it would be necessary to retrain the network on all transformations for every new object to be recognised.
- (ii) The demands placed on the classification system might be very severe in this approach. In a feature space of high dimension, the regions occupied by the transformed versions of an object will be arbitrarily distributed unless the feature space is suitable for such transformations requiring complicated decision boundaries. For optimal separation of these regions, the number of hyperplanes needed may be very large.

2.1.3 Recent approaches : the hybrids

In Section 2.1.1 we saw that it is possible to obtain transformation invariant features of an input image using moments. In Section 2.1.2 the abilities and limitations of a neural network classifier were discussed. In this section, we discuss the approach where invariant features are used as input to a neural network classifier [KHOTANZAD 90] [BARNARD 91].

There are two major advantages in this approach:

- (i) The requirements on the classifier are relaxed because the number of features required is much less than the number of pixels and the feature space is suited to the transformations.
- (ii) Invariance for all input objects is ensured.

The following are the disadvantages of using invariant feature spaces:

- (i) The input image is not directly input to the classifier. To calculate the features, preprocessing is required. Hence feature spaces must be **computationally** inexpensive to **calculate**.
- (ii) Not all feature spaces are equally suitable for a given problem. Each feature space has its own shortcomings. Moment invariant feature spaces

have difficulties when noise is present. Two other feature spaces, wedge ring samples of the magnitude of the Fourier transform, or the magnitude of the Fourier transform in log-polar coordinates are not invariant to all possible transformations.

2.1.4 Summary of Findings

- (i) It is possible to construct feature spaces invariant to simple metric transformations, namely, scaling, rotation and translation.
- (ii) We can construct effective classifiers only when a suitable invariant feature space is available. In other words, neural invariance by training is not practicable for realistic image sizes.
- (iii) Given an invariant feature space, a trained neural network classifier performs very well (% classification accuracy). It also exhibits desirable neural properties like learning capability, quick learning, fault tolerance and graceful degradation of performance.
- (iv) The moment invariant feature space has the advantages that it is simpler to compute, it works well for transformations over a wide range of scales and it needs less number of features for recognition.

2.2 LIMITATIONS OF APPROACHES IN THE CONTEXT OF DEGRADED IMAGES

Most of the studies reviewed above test the proposed techniques in simplified problem domains. Many aspects of complexity of the recognition problem as it may occur in a practical situation have not been considered. For example, **all** of them consider a single image **database**(e.g., the alphabet character set), whereas the performance of a recognition system may vary with the amount of details in the images. Similarly, in many studies in transformation invariant recognition, the issue of image degradation has not been addressed directly. However, some studies have considered [KHOTANZAD 90] [GRACE 91] degradation due to additive random noise, whereas in the context of imaging, as discussed in Chapter 1, degradation is of different nature.

Recognising objects from degraded images is inherently a complex task with many aspects of complexity. In such a situation, where various types of degradation occur in varying levels, it may not be reasonable to expect that a specific technique or method can be used with success always. Instead of diluting the problem to demonstrate the capabilities of a specific method, the problem must be addressed in all its complexity. Then various techniques and methods in hand can be applied on the problem and their applicability and limits can be studied. Hence the emphasis of the present study is to keep the issues in the problem domain constantly in view and study within what limits various approaches are useful.

2.3 PLAN OF THE PRESENT RESEARCH EFFORT

In this thesis, through simulation studies we create various cases where the **difficulty** of the task of recognition progressively increases and test systematically the performance of proposed **systems** which use neural network architectures and principles. As the complexity of the task is increased the range and limits of performance of the recognition systems are studied.

2.3.1 Overall Approach

Studies demonstrate the advantages of a system that employs neural network principles. Besides the advantages of adaptation these systems exhibit the desirable characteristics of (i) graceful degradation of performance with increased complexity and (ii) flexibility in application to different new contexts.

These **neural** network based methods can be used as standalone recognition **systems** in cases where the problem complexity is low. In more difficult cases they can aid the human decision maker.

2.3.2 Aspects Of Complexity

In this research effort we investigate the utility of neural network principles in developing recognition systems whose performance degrades gracefully when the problem becomes increasingly complex. For a systematic study we have chosen the

context of sensor array imaging.

The following aspects of complexity have been studied:

- (i) ***The goals of the recognition system*** : Classification or interpretation.
- (ii) ***Various types of degradation*** : Image distortions due to noise, incompleteness and poor resolution and image variability due to spatial transformations
- (iii) ***Varying levels of degradation***
- (iv) ***Complexity of the symbol set*** : Type and amount of details in the image and confusability among the symbols

2.3.3 Choice of Symbol Set

The type and amount of details in the images is another important factor that complicates the recognition task. If shapes are simple, then a system may exhibit successful recognition performance over a wide range of distortions. This, however, does not guarantee that the system can be applied to images of arbitrary complexity. Hence we have deliberately chosen a set of symbols with sufficient details and confusability so that the level of degradation with which recognition can be performed can be studied. These methods give much better performance when the objects are simpler in terms of detail.

2.3.4 Study in three stages

Studies in object recognition are conducted in three different cases. In the first case we consider the issues of image degradation alone. In the second case we consider the issue of transformational variability alone. In the third case we consider the issues together. This is done in order to understand the difficulties in addressing each issue in isolation. Another outcome of such a study done in stages is that it allows us to evaluate performance degradation as the complexity increases.

In the following chapters we describe these studies.

3

OBJECT RECOGNITION FROM DEGRADED IMAGES

In this chapter we consider the situation where images are poorly resolved and corrupted by noise. In such situations where feature extraction may not be successful, we aim to exploit the noise immunity and learning capabilities of a neural network for object recognition. We describe the approach of correlation matching using a neural network in Section 3.2. Section 3.3 describes Hamming network used for recognition. In Section 3.4 various experimental studies in object recognition from images reconstructed from sparse and noisy sensor array data are described. Results are discussed in section 3.5.

3.1 DIFFICULTIES IN OBJECT RECOGNITION FROM DEGRADED IMAGES

In the case of sensor array imaging situation, noise and sparsity of data is in a **transform domain** and hence all pixels are affected rather than individual pixels of the image. Feature matching algorithms that use line and curve extractors cannot be used directly in this case as they may detect spurious lines and arcs. The image may be preprocessed before feature extraction to remove the noise. This approach is taken in Chapter 5. However, such preprocessing techniques may be applied only **upto** certain levels of degradation, beyond which their output may be unreliable.

If the goal is just classification and not generation of a description of the image, one may not need preprocessing and feature extraction stages. Instead, a pixelwise

template matching approach may be preferred over feature based methods in such situations. In this chapter we describe a simple neural network that learns templates of the images and matches by correlation.

3.2 CORRELATION MATCHING USING A NEURAL NETWORK

We use Hamming network [LIPPMAN 87] which performs template matching using neural principles. A neural network model has the following advantages in performing template matching over a nonneural implementation:

1. Templates can be learnt from examples instead of explicit specification.
2. The degree of match with several templates can be calculated in parallel, owing to their parallel structure. Thus pixel based correlation can be easily implemented using these models.
3. The best match can be found by competition between various hypothesis instead of a simple matching score.
4. Networks exhibit some noise immunity by their capability to adaptively threshold the individual neuronal activity.

3.3 THE HAMMING NETWORK

3.3.1 Structure

The Hamming net is a maximum likelihood classifier for binary inputs corrupted by noise [LIPPMAN 87]. It consists of two **subnets**, the lower and the upper ones as shown in Fig.3.1. The lower **subnet** consists of N input nodes, each corresponding to a pixel in the pattern and M output nodes corresponding to the M patterns in the knowledge base. The upper **subnet** consists of M nodes connected to each other through inhibitive weights.

3.3.2 Learning

Initially weights are learnt **from** examples:

In the lower **subnet**, the connection weights between the input and output nodes

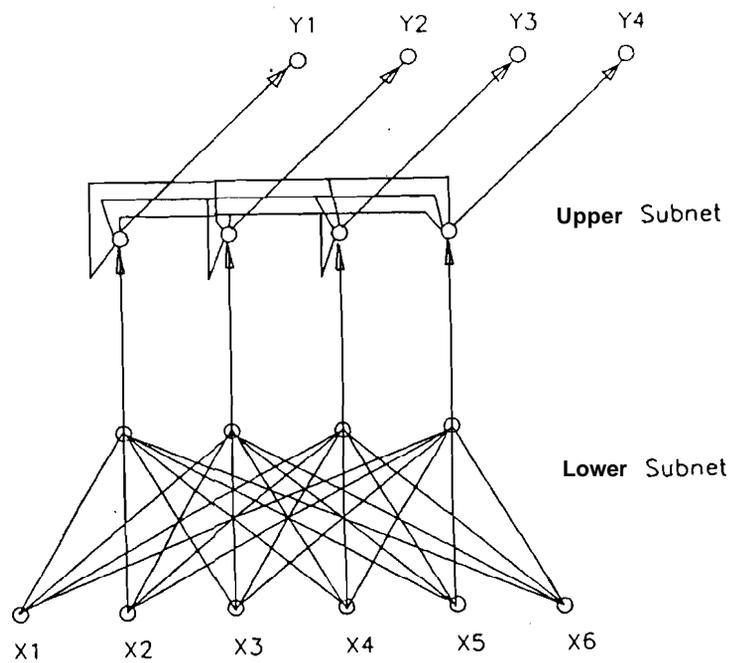


Fig.3.1 The Hamming network. X's represent the input nodes and Y's represent the output nodes.

are fixed in such a way that the network calculates the distance from the input noisy pattern to each of these M pattern classes:

$$w_{ij} = x_{ij} / 2, \quad \Theta_j = N/2$$

$$0 < i < N-1, \quad 0 < j < M-1$$

where w_{ij} is the connection weight from input node i to output node j in the lower subnet,

Θ_j is the threshold at that node and

x_{ij} is the element i of the pattern j .

In the **upper subnet**, weights are fixed in such a way that the output nodes inhibit each other:

$$t_{kl} = 1, \quad k = l$$

$$= -\epsilon, \quad k \neq l$$

where t_{kl} is the connection weight in the upper net from node k to node l .

3.3.3 Classification

When a binary pattern is presented for classification, the lower **subnet** calculates its matching score m_j with each pattern in the knowledge base:

$$m_j(0) = f(\sum_i w_{ij} x_i - \Theta_j),$$

where $m_j(t)$ is the output of node j in the upper **subnet** at time t , x_i is the element i of the input pattern and $f(x)$ is the threshold function.

This is presented to the upper **subnet** and competitive interaction takes place among the output nodes:

$$m_j(t + 1) = f(m_j(t) - \varepsilon \sum_{k, k \neq j} m_k)$$

Competition continues until the output of only one node becomes positive. This selected node corresponds to the recognised pattern.

3.4 EXPERIMENTAL STUDIES

The primary aim of these studies is to test the performance of the network under different levels of degradation and to **identify** the limits **upto** which the network can **perform** reliably. Sparsity of samples and noise are the main reasons for the poor quality of the reconstructed image. Hence to study the recognition performance, three different experiments were performed. Images were reconstructed from sparse data collected **using** arrays of different sizes in Experiment 1, data with various levels of noise in Experiment 2 and data collected at multiple frequencies in Experiment 3.

The knowledge base in the present studies consists of 20 **olympic** games symbols shown in **Fig.3.2**. These symbols were created by first **digitising** the printed symbols from a newspaper **using** a scanner and editing the symbols. Each image consists of 128x128 pixels.

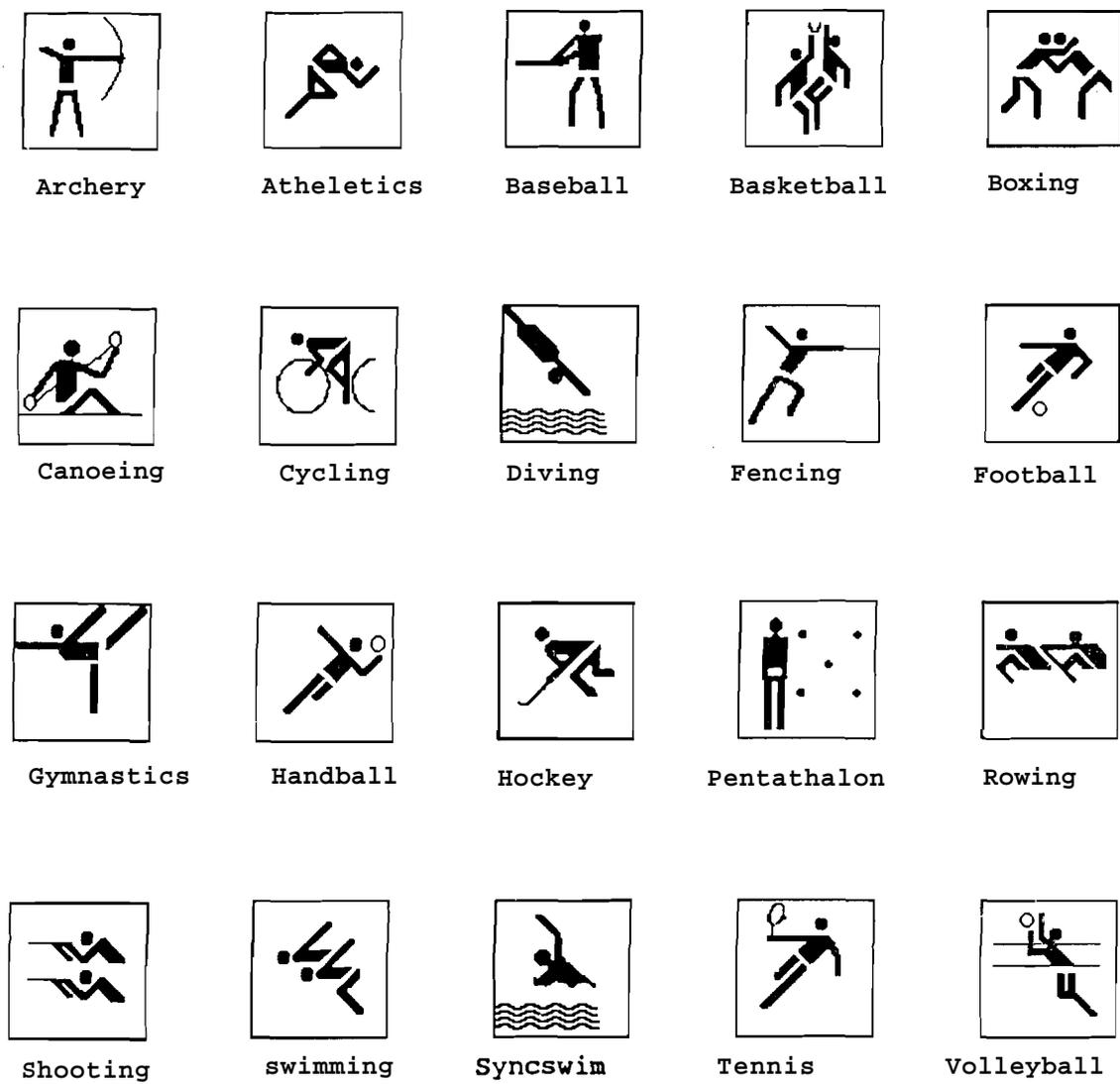


Fig.3.2 Olympic game symbols (20) in the knowledge base used for the studies in recognition of objects from degraded images.

3.4.1 Effect of increasing sparsity

Data at the sensor array was simulated assuming array-sizes of 64x64, 32x32, 16x16 and 8x8 sensors, using in each case two frequencies for data collection [YEGNANARAYANA 91]. Fig.3.3 shows the images reconstructed from the 16x16 array data and Fig.3.4 from 8x8 array data. Each of these images was converted into a binary image and presented to neural network which has been trained to the 20 original images. The values given in the brackets are the activation values of the winning node.

The results are summarised in Fig.3.5. All the twenty patterns are correctly classified when the array size is large than 8x8. This performance is very impressive since it is difficult for us to identify visually discriminating features in many of these images.

3.4.2 Performance with data collection at multiple frequencies

Sensor array data was simulated assuming an 8x8 array with data collected at different frequencies. The images reconstructed from two and four different frequencies data, shown in Fig3.4 and Fig.3.6 respectively, were presented to the network. The results are **summarised** in Fig.3.7. Classification performance is accurate when four or more frequencies are used for data collection if the array size is 8x8.

3.4.3 Effect of noise

Random noise with a Gaussian distribution was added to the simulated sensor array data collected at two frequencies by a 16x16 array. Images reconstructed from this noisy data are shown in **Fig.3.8** and Fig.3.9, corresponding to an **SNR** of -3dB and -6dB, respectively. When noisy images were presented to the network, the results obtained are **summarised** in Fig3.10. It appears that noise causes a kind of degradation different from and more severe than the degradation caused sparsity of data. However, for the case of 16x16 data, the network classifies satisfactorily **upto** a noise level of -3dB.

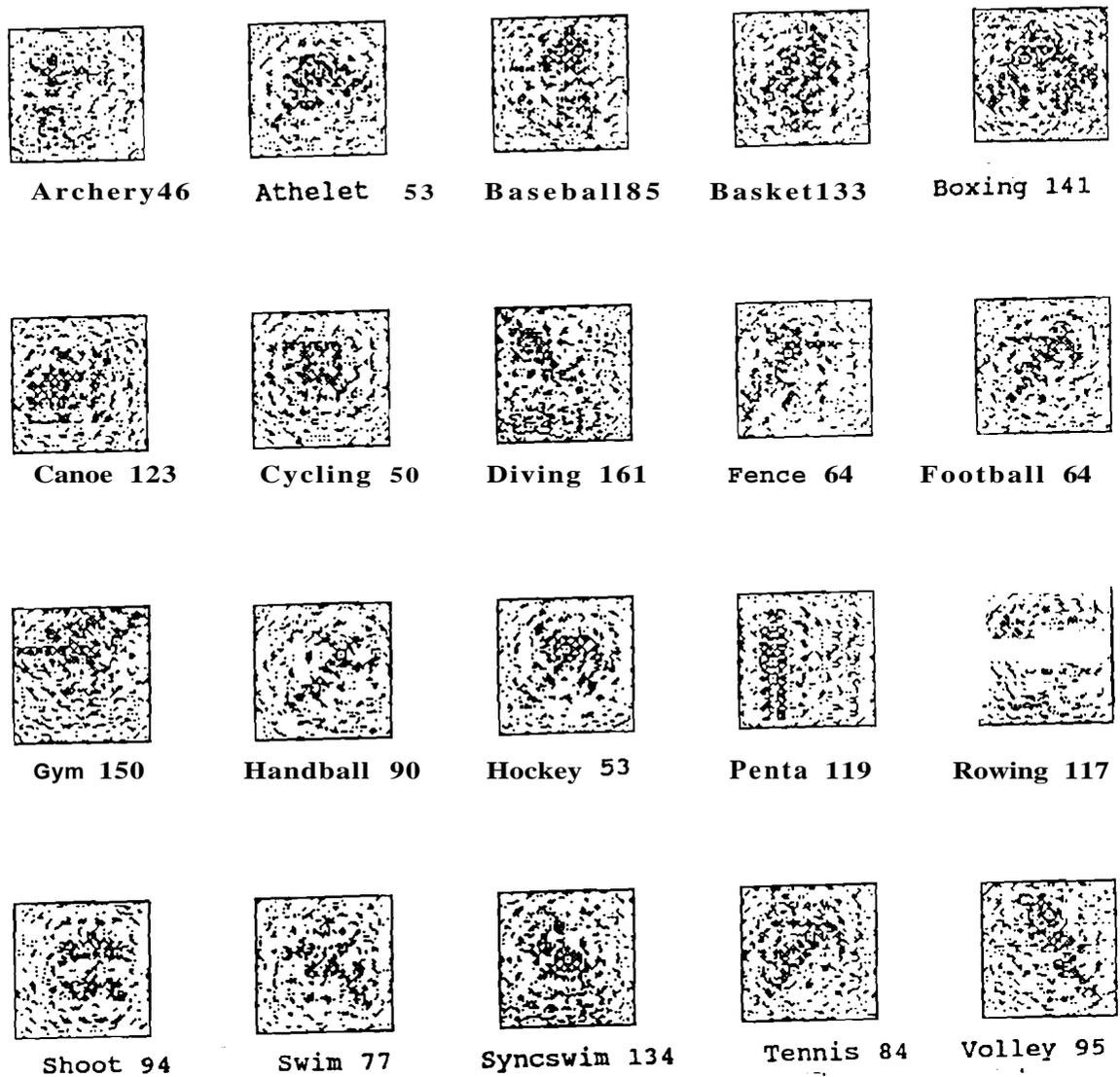


Fig.3.3 Recognition results for images reconstructed from data collected at two frequencies from a 16x16 array. The class decision of the network is given along with the activation value of the winning pattern. In this case, all of the twenty images were correctly identified.

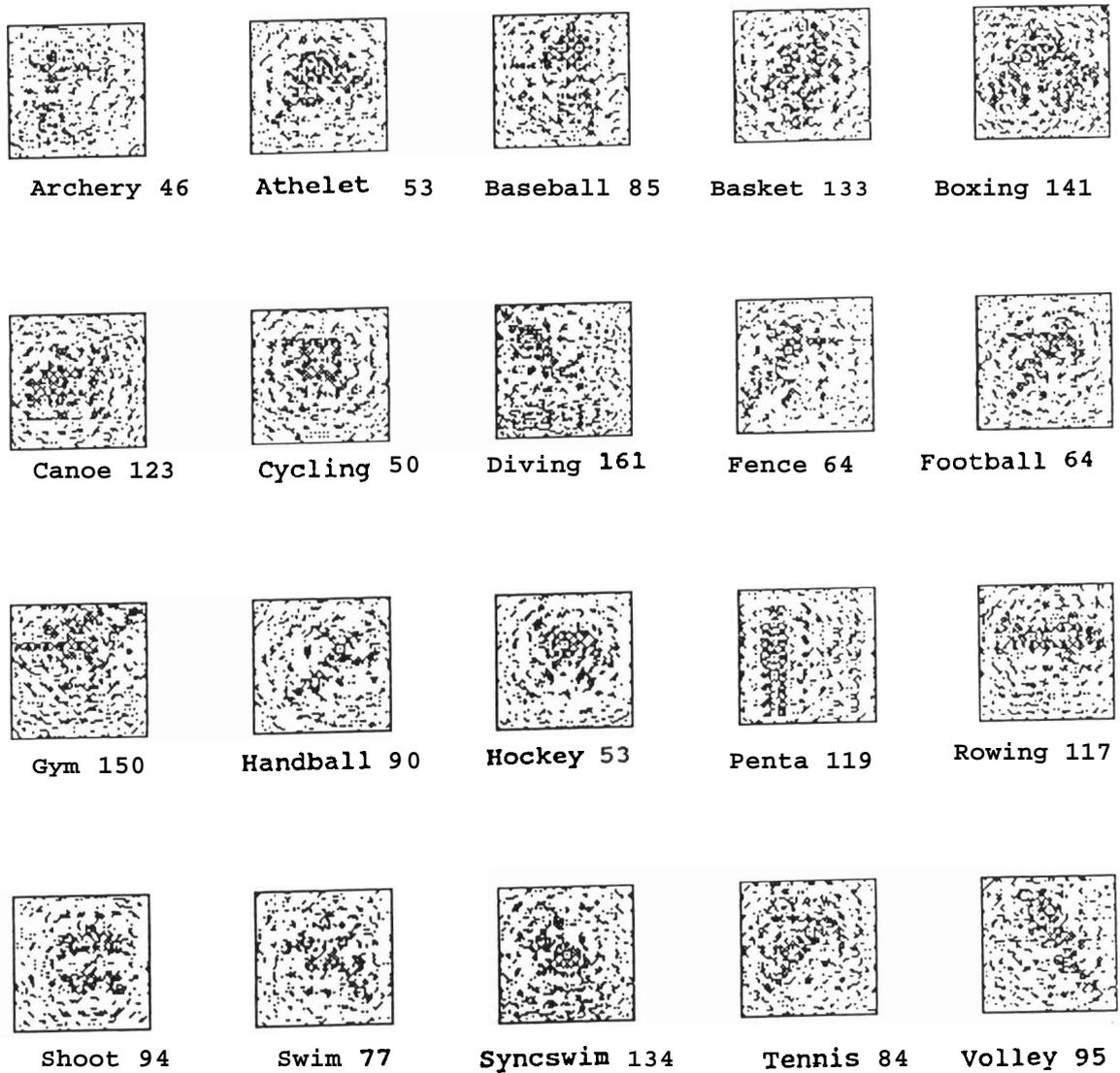


Fig.3.3 Recognition results for images reconstructed from data collected at two frequencies from a 16x16 array. The class decision of the network is given along with the activation value of the winning pattern. In this case, all of the twenty images were correctly identified.

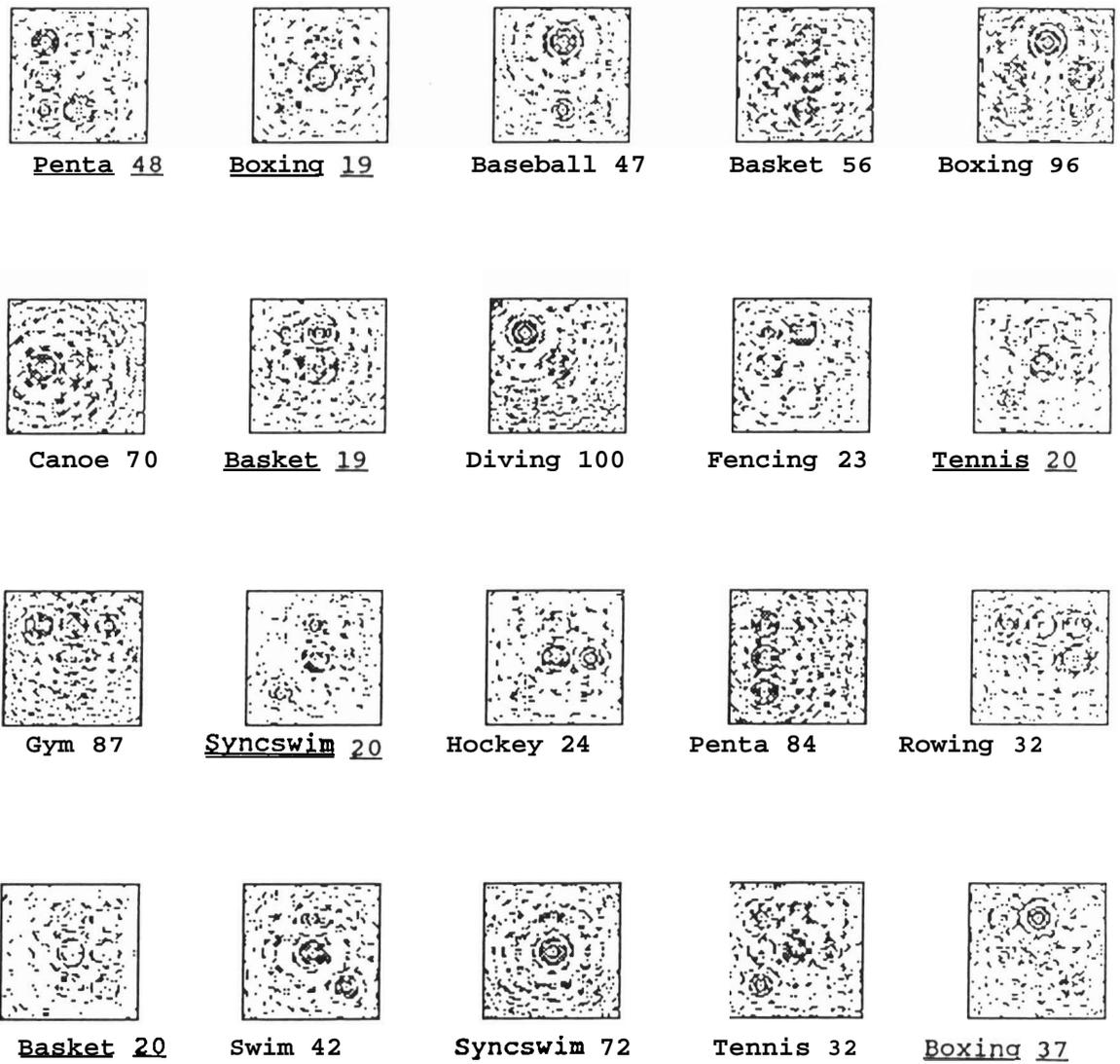


Fig.3.4 Recognition results for images reconstructed from data collected at two frequencies from an 8x8 array. The class decision of the network is given along with the activation value of the winning pattern.

Number of images correctly identified = 13.

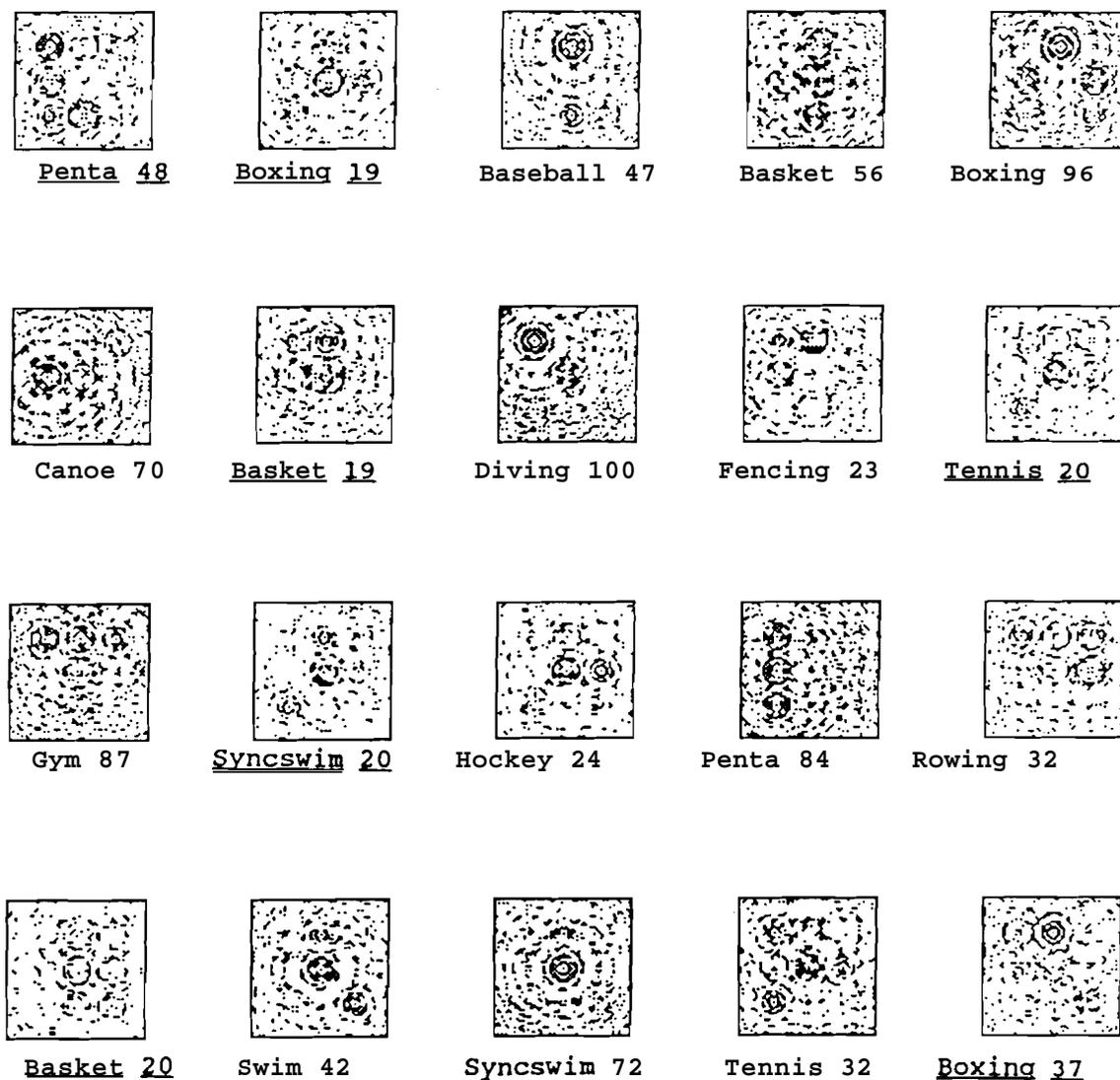


Fig.3.4 Recognition results for images reconstructed from data collected at two frequencies from an 8x8 array. The class decision of the network is given along with the activation value of the winning pattern.

Number of images correctly identified = 13.

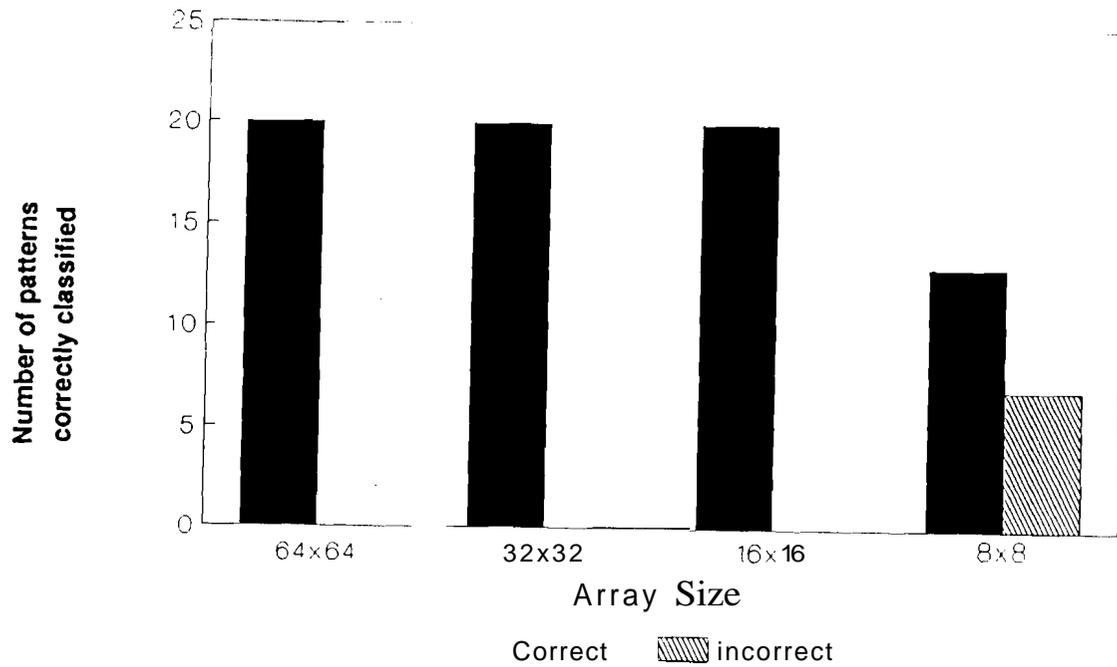


Fig.3.5 Summary of recognition performance with different sparse arrays (64x64, 32x32, 16x16 and 8x8 sensors). Graph shows the number of patterns correctly identified out of twenty patterns in each case.

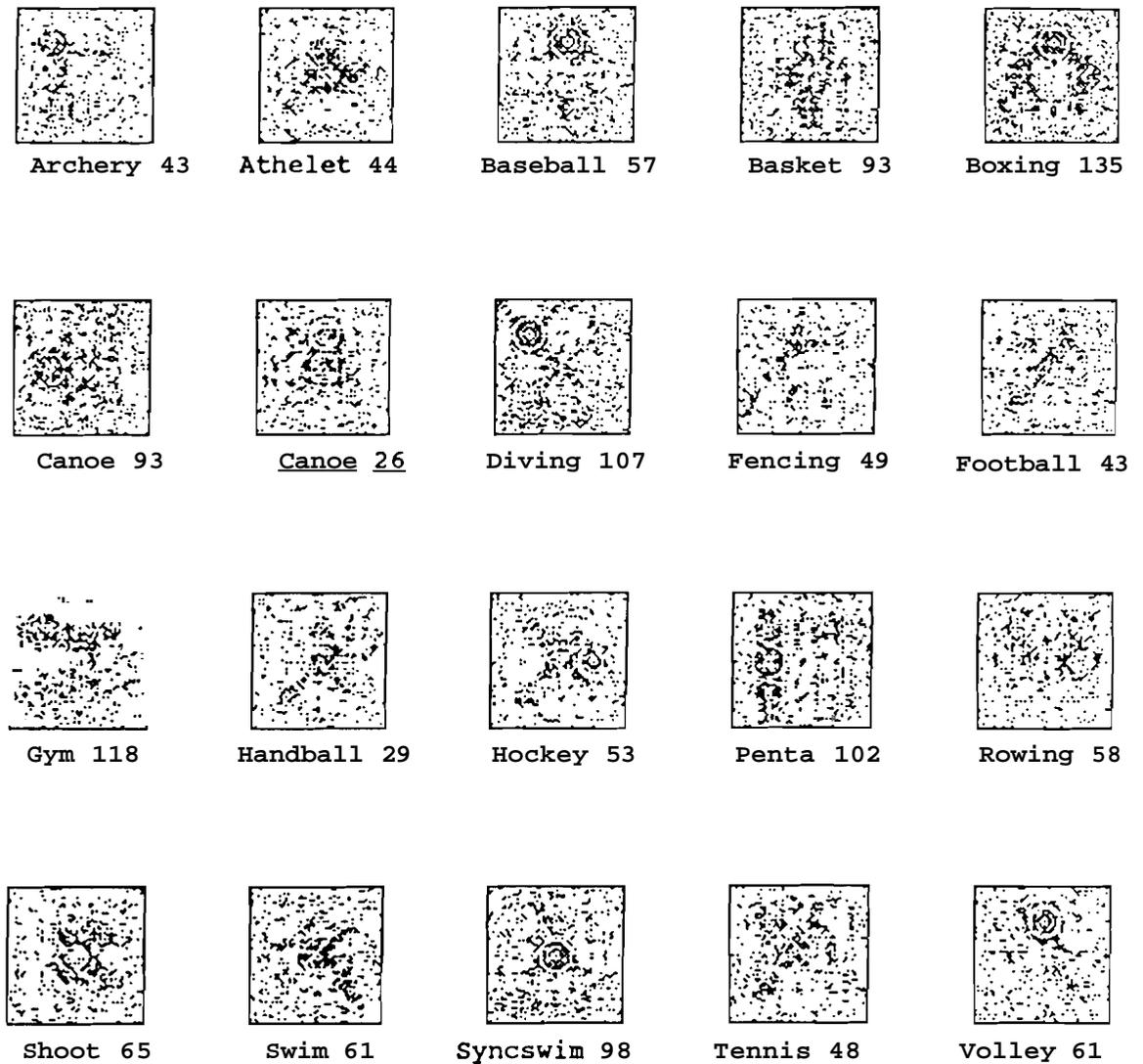


Fig.3.6 Recognition results for images reconstructed from data collected at four frequencies from an 8x8 array. The class decision of the network is given along with the activation value of the winning pattern. Misclassification is underlined.

Number of images correctly identified = 19.

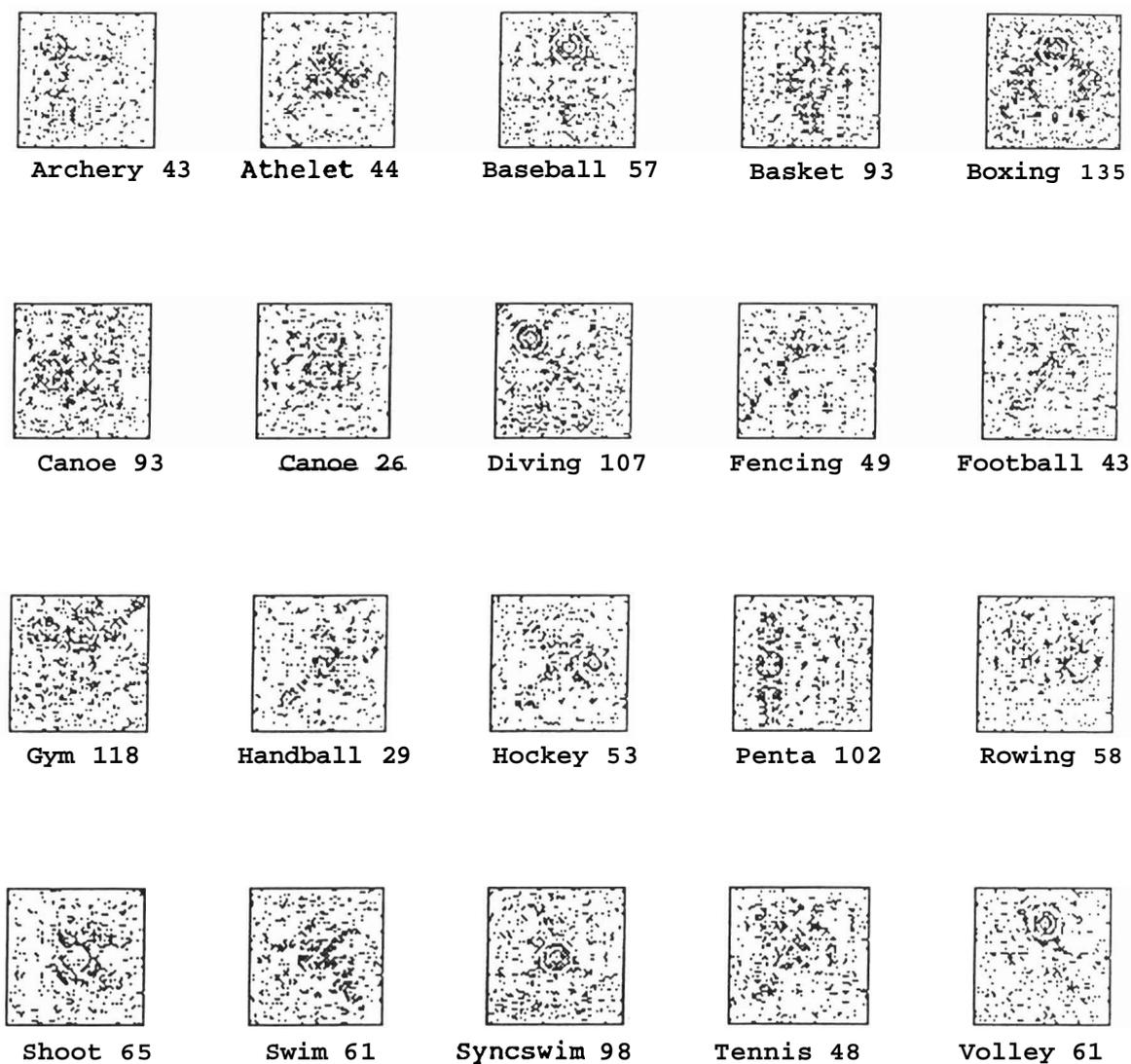


Fig.3.6 Recognition results for images reconstructed from data collected at four frequencies from an 8x8 array. The class decision of the network is given along with the activation value of the winning pattern. Misclassification is underlined.

Number of images correctly identified = 19.

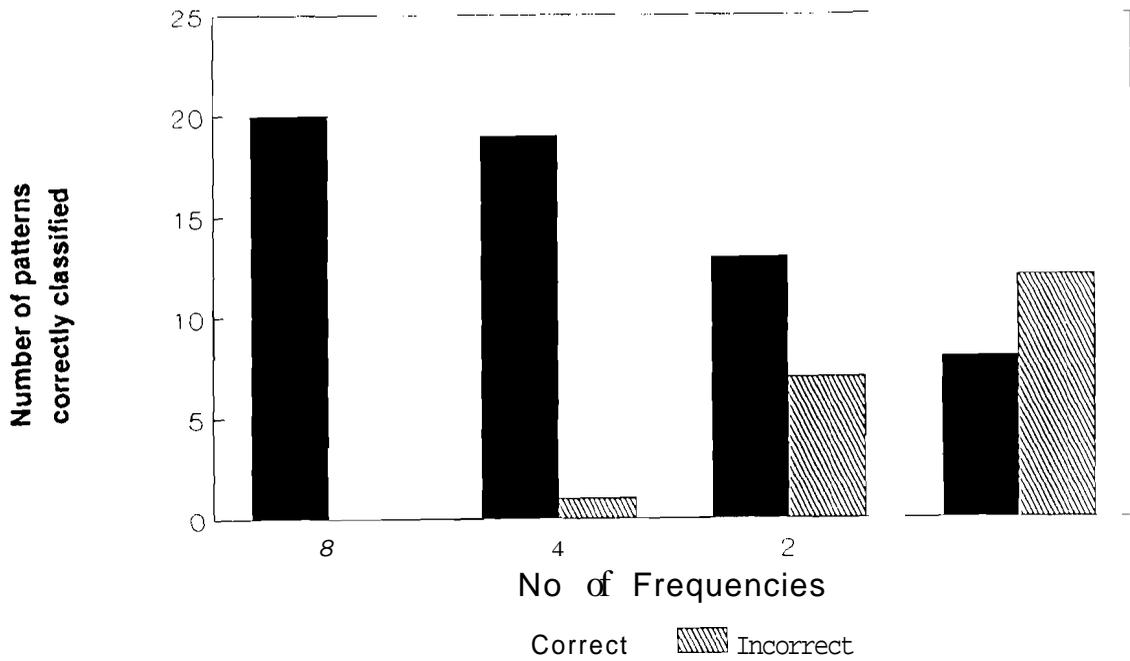


Fig.3.7 Summary of recognition performance with different cases of multiple frequency data collection, for 8, 4, 2 and 1 frequencies used for data collection using an 8x8 sensor array. Graph shows number of patterns correctly identified out of twenty patterns in each case.

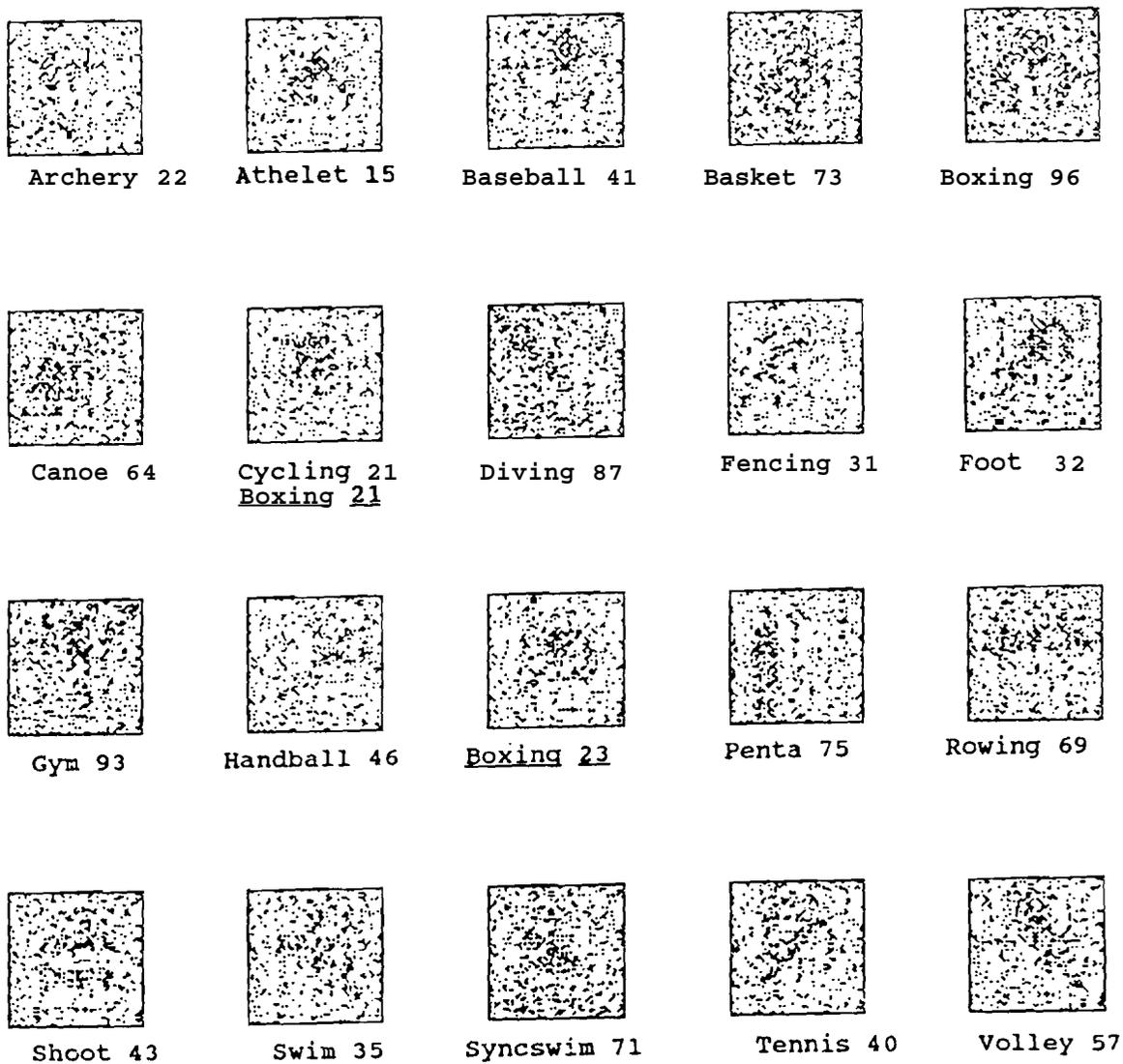


Fig.3.8 Recognition results for images reconstructed from noisy data (SNR = -3dB) collected at two frequencies from a 16x16 array. Noise was Gaussian and added to the received data. The class decision of the network is given along with the activation value of the winning pattern. Misclassification is underlined.

Number of images correctly identified = 18.

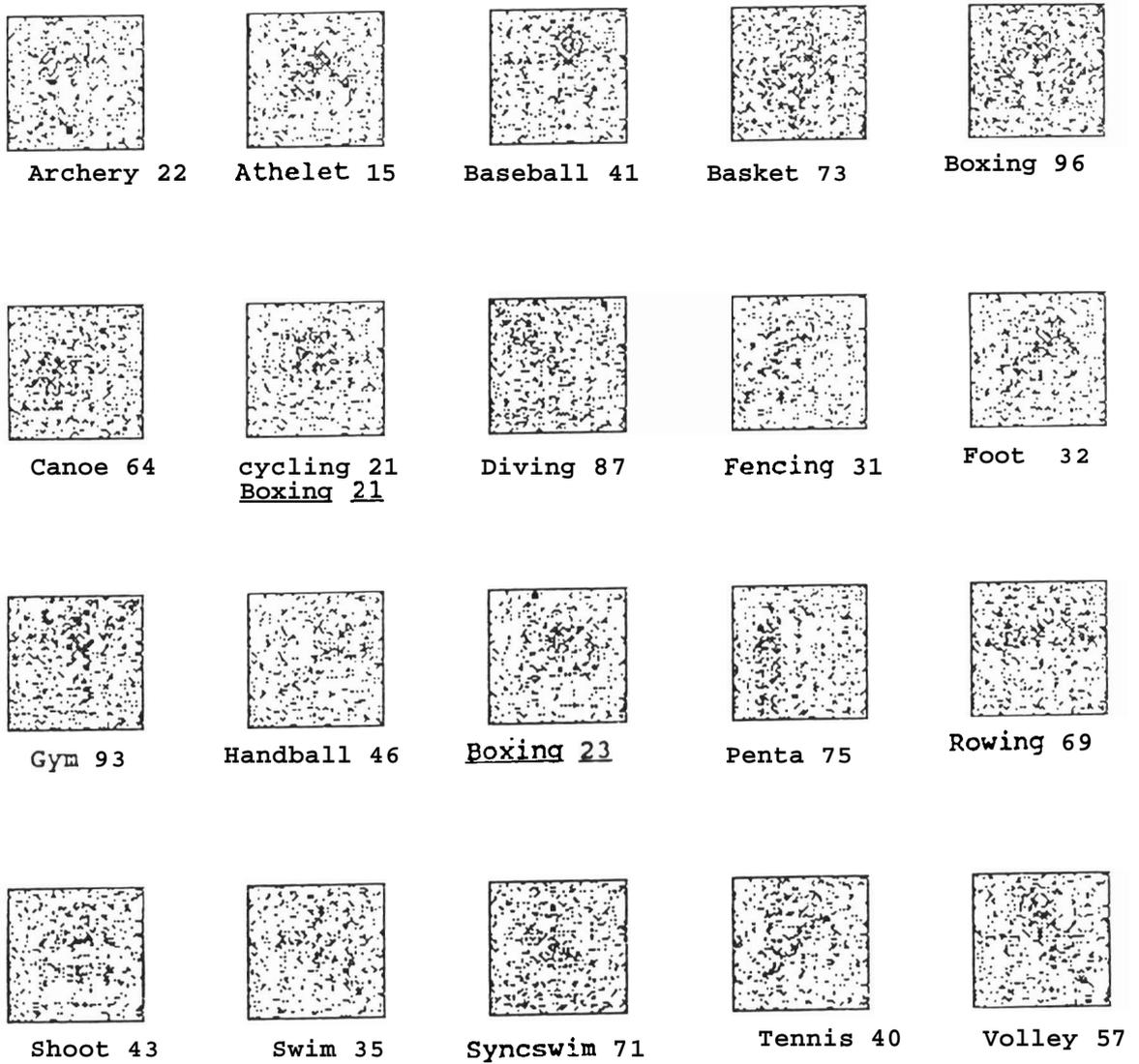


Fig.3.8 Recognition results for images reconstructed from noisy data (SNR = -3dB) collected at two frequencies from a 16x16 array Noise was Gaussian and added to the received data. The class decision of the network is given along with the activation value of the winning pattern. **Misclassification** is underlined.

Number of images correctly identified = 18.

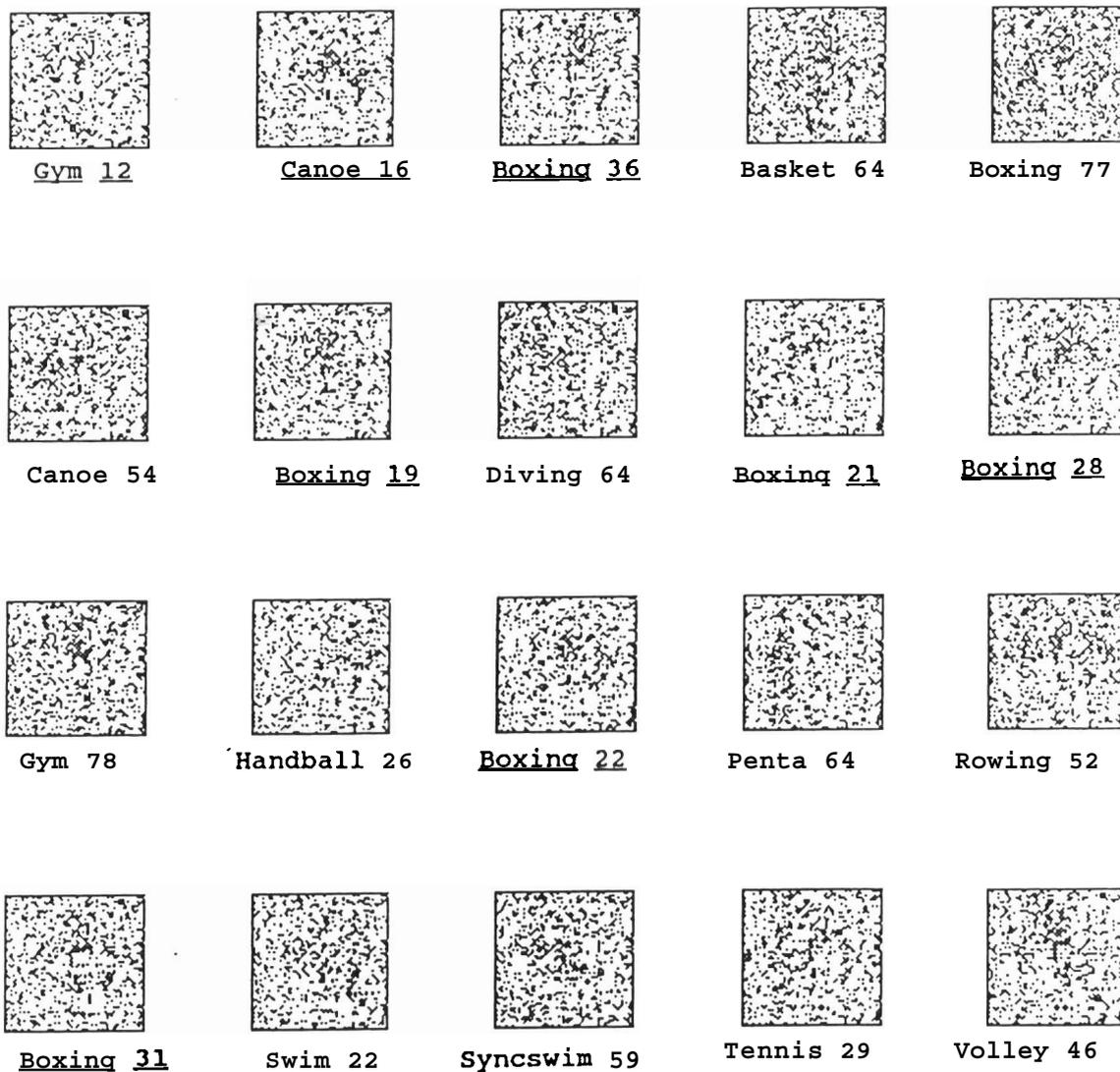


Fig.3.9 Recognition results for images reconstructed from noisy data (SNR = - 6dB) collected at two frequencies from a 16x16 array. Noise was Gaussian and added to the received data. The class decision of the network is given along with the activation value of the winning pattern. Misclassification is underlined.

Number of images correctly identified = 12.

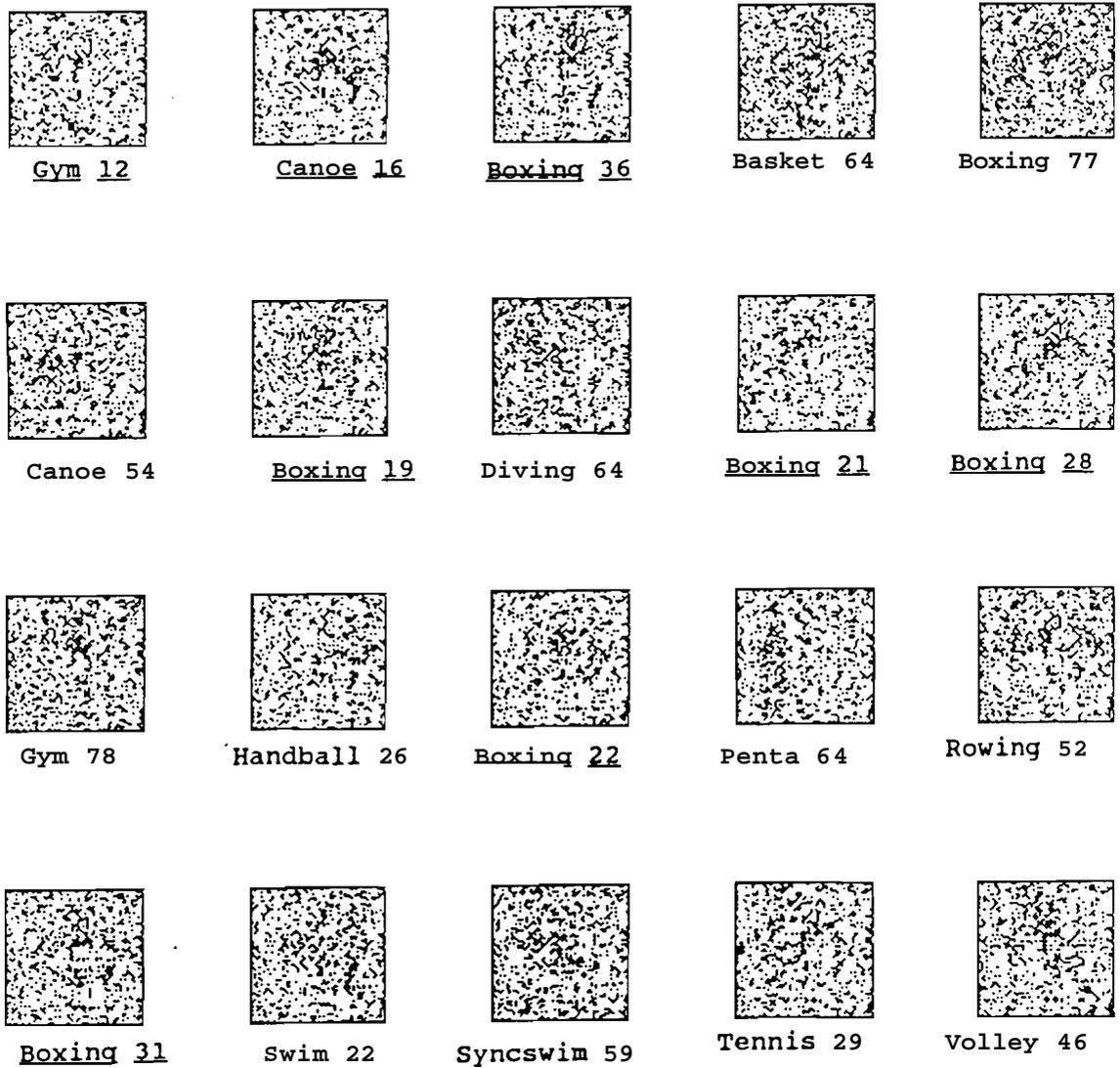


Fig.3.9 Recognition results for images reconstructed from noisy data (SNR = - 6dB) collected at two frequencies from a 16x16 array Noise was Gaussian and added to the received data. The class decision of the network is given along with the activation value of the winning pattern. Misclassification is underlined.

Number of images correctly identified = 12.

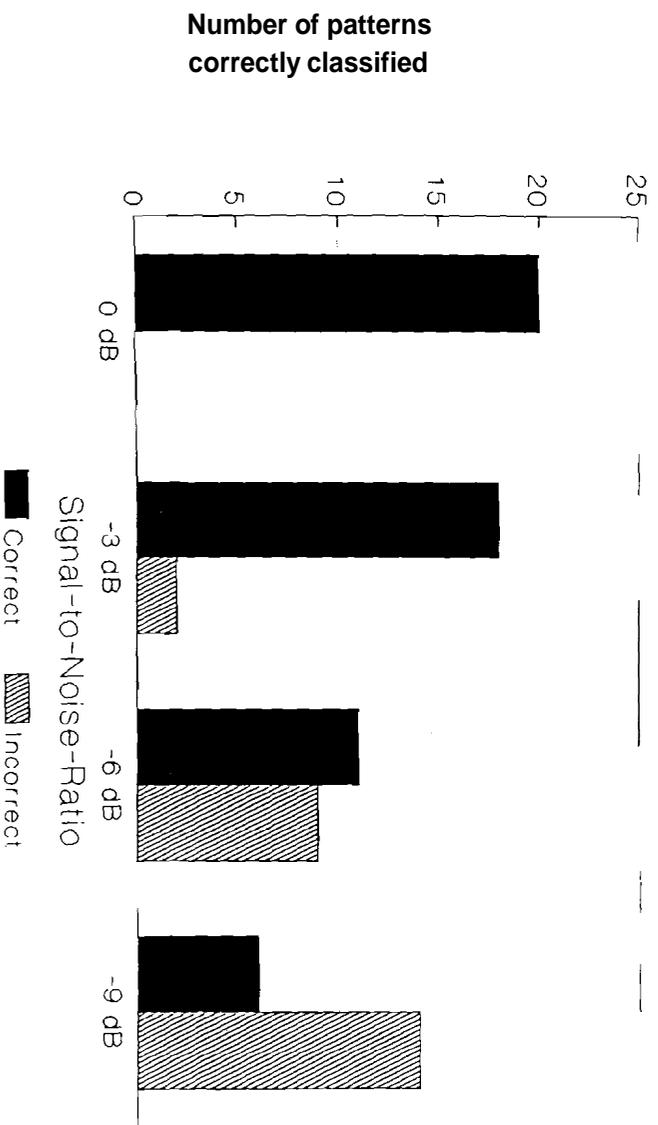


Fig. 3.10 Summary of recognition performance with different levels of noise in the received data collected at two frequencies by a 16x16 array. Graph shows number of patterns correctly identified out of twenty patterns in each case.

3.5 RESULTS AND DISCUSSION

- (i) It is interesting to note that many images in the 16x16 case which seem to have very few visual clues for us to recognise are recognised correctly by the network. This shows that the performance of a machine may be superior to human performance when the situation is unusual for a human being. The machine may perform much better by using techniques quite different from what a human being uses, even though these techniques may seem to be simple.
- (ii) The performance of the classifier degrades gradually with increasing image degradation. With increasing sparsity and noise, classification accuracy drops gradually as seen in each of the graphs in Figs. 3.5, 3.7 and 3.10.
- (iii) The activation values of the winner node are indicative of the level of image degradation. Greater the degradation, lower is the activation of the winning nodes. For example, the value reduces consistently with increasing **degradation** for any given pattern, as seen in Figs. 3.3 and 3.4 (images reconstructed using **16x16** and 8x8 array data respectively).
- (iv) The matching scores obtained at the first layer are measures of similarity of the input pattern with each pattern in the knowledgebase. However in the second layer, the activation level of each node is affected by the activation of every other node. Hence when an output node becomes positive, its activation level not only reflects how close the input image is to the identified pattern, but also the amount of confidence given to this decision with respect to other patterns in the knowledgebase. Thus the value also indicates the complexity of the symbol set, for example, in terms of how close in shape the symbols are.
- (v) Subjective judgements regarding the quality of the images appear to be confirmed by the activation of winning nodes. Perceptually, for example, the images reconstructed from data collected by an 8x8 array at 4 frequencies (Fig.3.6) are poorer in quality than images reconstructed from data collected by a 16x16 array at 2 frequencies (Fig.3.3). The

winning scores in the latter case are higher than those in the former, as can be seen from these figures.

3.6 SUMMARY

Major results : If the set of expected objects are known, then it is possible to train a neural network for object recognition. Even though the present network performs a simple correlation matching, it classifies even images which are so degraded that we fail to perceive discriminating features visually. Thus this study demonstrates that in some situations, a trained neural network classifier can perform object recognition in the presence of severe noise and sparsity.

Limitations: In the study since it was assumed that the issues of transformation and spatial distortion do not exist, the network used direct pixel-wise description. Hence its application is limited to simple situations. In practice this condition is seldom satisfied since objects being imaged move relative to the imaging system. Fig.3.11 shows some images reconstructed from simulated sensor array data when the object moved with respect to the array, resulting in spatial transformations of the image. It is extremely difficult to identify objects from these images using the Hamming net.

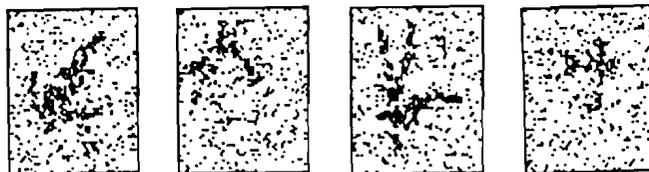


Fig.3.11 Images reconstructed from sensor array data collected when the objects have moved relative to the array.

If images are clean and noise-free, there exist methods for overcoming the effects of metric transformations. In the following chapter, we explore this possibility, *i.e.*, transformation invariant recognition of objects from noise-free images before attempting it in the case of noisy images.

4

TRANSFORMATION INVARIANT RECOGNITION OF OBJECTS

In this chapter we consider the issue of transformation, but assume a noise free situation. We study the capability of a neural network to learn object shapes from examples and generalise its learning to accomodate variability.

We discuss the issues in transformation invariant recognition of objects in Section 4.1. Approaches described in literature are briefly reviewed and our approach is discussed in Section 4.2. Section 4.3 describes the derivation of a transformation invariant feature space based on the theory of geometric moments. Section 4.4 describes a neural network classifier which uses this feature space for object recognition. This is followed by the description of experimental studies in Section 4.5 and discussion of results in Section 4.6.

4.1 THE PROBLEM OF TRANSFORMATIONAL VARIABILITY

As an object moves around in its environment the image obtained from any imaging system may be a translated, rotated and scaled version of its template in the knowledge base. Hence for recognition of objects, the image or its description must be **normalised** with respect to these spatial transformations before matching is performed.

In simplified situations, for example the class of tasks that involve strictly: unoccluded, segmented objects, one may not require descriptions based on local features followed by a separate step of normalisation. Instead global matching techniques can be used in which measures invariant to the transformations of interest

are computed directly from the image and used as descriptors.

4.2 APPROACHES TO SOLVE THE SPATIAL TRANSFORMATIONS PROBLEM

Invariant object recognition can be approached in two different ways:

4.2.1 Single stage neural network approaches

Several neural network models [BARNARD 91] for invariant pattern recognition have been proposed. In the case of networks which achieve *invariance by structure*, the structure of the network is designed such that the output is always invariant to the transformations of interest. In the second case of *invariance by training*, representatives of large class of transformations are presented during training so that the network learns equivalent transformations.

4.2.2 Invariant feature approach

Alternatively, recognition may be done in two stages. The images are first normalised with respect to size, position and orientation by estimating the transformation parameters. This is followed by matching using either templates or feature descriptions. Equivalently invariant feature measures may be computed directly **from** the transformed image and be used for classification.

Theoretically, we can build a fully connected feedforward multilayer perceptron network and train it by error backpropagation in such a way that it can perform successful object recognition. The size of the network required for this task may be very large. Moreover, such an **unconstrained** superfluous structure may result in the network blindly memorizing the input-output relationships without attempting to **generalise**. Such a network cannot extend its knowledge to handle new examples of the same input pattern. Generalisation also depends on the type of training examples. The question of how to constrain the structure to force the network to generalise and how to select and sequence the training samples so that the generalisation will be a valid one, are questions yet to be answered. In such a situation, a practical approach would be to handle the problem of normalisation separately so that the classification task

becomes relatively simpler.

Invariance by structure or training presupposes the existence of a fixed set of weights which can provide invariance over the continuum of transformations. It also assumes that a network can be trained to estimate this set of weights from examples. We argue that invariances cannot be built as a static function of structure, but have to be dynamically estimated from the data. Hence our approach is to handle the issue of transformational variability first and then use a classifier [RAVICHANDRAN 91].

4.3 THE INVARIANT FEATURE SPACE

Feature extraction, in general, is the process in which the image is represented by a set of numerical features. This is done for two reasons:

- (i) ***To reduce the dimensionality of the input pattern:*** If pixels have totally uncorrelated and random distributions in all the input images, then features **cannot** be seen. However, in the image of any object, the constituent pixels are related to their neighbourhoods and this makes a large number of pixel distributions meaningless. This regularity and the resulting redundancy can be exploited. Pixel distributions that occur often can be named as features. With a suitable set of such features, the image can be described in terms of these features leading to reduced dimensionality.
- (ii) ***To allow for variations and distortions:*** By abstracting away the pixel distributions **from** their exact physical locations, features provide for invariant description of objects. When an object undergoes spatial transformation, the spatial positions vary, but certain spatial interrelationships between points on the object may be maintained. For example, after translation or rotation the positions of the individual pixels change, but their relative distances remain unaltered. Even after scaling the ratios of relative distances remain same.

Where normalisation is not required for classification, it is possible to define measures that are invariant to spatial transformations and represent the image by a set of numerical features. In this section, we describe such an invariant feature space.

4.3.1 Theory of moments

Methods based on the theory of geometric moments have been used for normalisation and invariant feature extraction [HU 62]. If the object is compact and has few details, invariant measures stable over a wide range of spatial transformations can be designed. If the transformations are metric, then it is possible to design such an invariant feature space.

In general, moments are numerical quantities which describe a distribution. In statistics moments are used to characterise the distribution of a random variable and in mechanics to characterise bodies by the spatial distribution of their mass. If we consider a binary or a grey valued image segment as a two-dimensional density distribution function, then moments may be used to characterise an image segment.

Given a two dimensional $M \times M$ image $\{ f(x,y); x,y = 0,\dots,M-1 \}$, the $(p+q)$ th geometric moment is defined as

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} x^p y^q f(x,y)$$

for $p,q = 0,1,2, \dots$

Note that for this moment definition, the monomial product $x^p y^q$ is the **basis function**. To keep the dynamic range of m_{pq} consistent for different size images, the $M \times M$ image plane is first mapped on to a square defined by $x \in [-1, +1]$, $y \in [-1, +1]$. **Hence** grid locations will have real values in the $[-1, +1]$ range. This changes the definition of m_{pq} to

$$m_{pq} = \sum_{x=-1}^{+1} \sum_{y=-1}^{+1} x^p y^q f(x,y).$$

To make these moments translation invariant one can define a central moment as

$$\mu_{pq} = \sum_{x=-1}^{+1} \sum_{y=-1}^{+1} (x-\bar{x})^p (y-\bar{y})^q f(x,y).$$

with $\bar{x} = \frac{m_{10}}{m_{00}}$ and $\bar{y} = \frac{m_{01}}{m_{00}}$

Central moments can be normalised to become invariant to scale change by defining

$$\eta_{pq} = \mu_{pq}^\gamma, \text{ where } \gamma = \frac{(p+q)}{2} + 1.$$

The lower order moment values represent well known fundamental geometric properties of a distribution. Consider a distribution function that is binary and **contiguous** such as a silhouette **image** of a segmented object. The moment values of this distribution may be explained in terms of simple shape characteristics of the object. The zeroth order moment **m00** represents the total object **area**. The two first order moments **m10** and **m01** are used to locate the center of **mass** of the object. The centre of mass defines a unique location with respect to the object which may be used as a reference point to describe the position of the object with the field of view.

The second order moments **m02**, **m11** and **m20** known as moments of inertia, may be used to compute useful **object** features like, the **principal axes** of the distribution, the image ellipse, which is a constant intensity elliptical disk with the same mass and second order moments as the original image and the same radii of gyration.

4.3.2 Algebraic moment-invariants

The use of moments for **image** analysis and object representation was inspired by Hu [HU 62]. Based on theories of invariant algebra that deal with properties of algebraic expressions which remain invariant under general linear transformations, Hu derived combinations of moment values that are invariant with respect to scale,

position and orientation. They are

$$\varphi_1 = \eta_{20} + \eta_{02}$$

$$\varphi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$\varphi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$\varphi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\begin{aligned} \varphi_5 = & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \cdot [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ & + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \cdot [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned}$$

$$\varphi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{03} + \eta_{21}).$$

These numerical values are very small, but vary over a wide range. To avoid problems of precision, the logarithms of the absolute values of these six functions are selected as features representing the image.

The utility of the moment invariants is illustrated through the following experiment. Fig.4.1 shows several olympic game symbols represented in a two dimensional feature space formed by the first two moment invariants. Points representing each of the symbols are distinct.

It must be noted that some of the symbols that are very different in image shape are close to each other in the feature space. Hence moment invariant features may not correspond to visually discriminating features employed by the human visual system.

4.4 THE CLASSIFIER

4.4.1 Classification

Classification consists of associating the feature vectors with the corresponding output symbols. This consists of a learning phase in which the system is taught the input-output relationships. Based on this knowledge, the system classifies the input patterns.

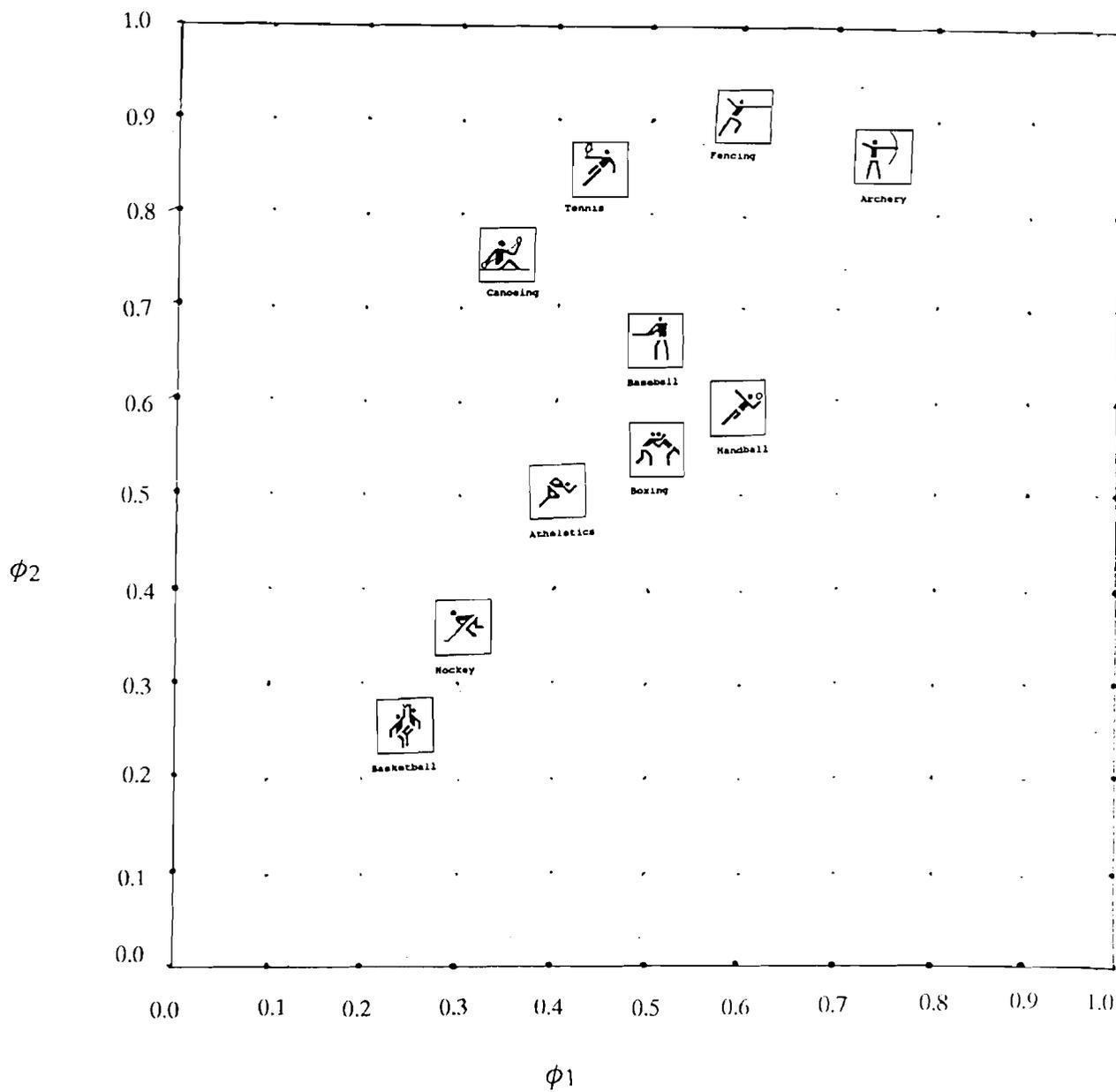


Fig.4.1 Several Olympic game symbols represented in a two dimensional feature space formed by the first two moment invariant features, ϕ_1 and ϕ_2 . Values represent logarithm of the absolute values of the features, normalised to unity.

To perform classification based on these features, we use a multilayered neural network trained using the error backpropagation algorithm [RUMELHART 86B].

4.4.2 Multilayer neural networks

Multilayer feedforward networks have been used as powerful classifiers. Their superiority over nonneural classifiers in forming arbitrary decision regions in multidimensional vector spaces has been demonstrated. In this process they also generalise better than conventional techniques [RUMELHART 86B].

In this study, a multilayer **perceptron(MLP)** is used for classification. An MLP is a **feedforward** network with one or more layers of nodes between the input and the output nodes. These in-between layers are called hidden layers. An **MLP** with one hidden layer is shown in Fig.4.2.

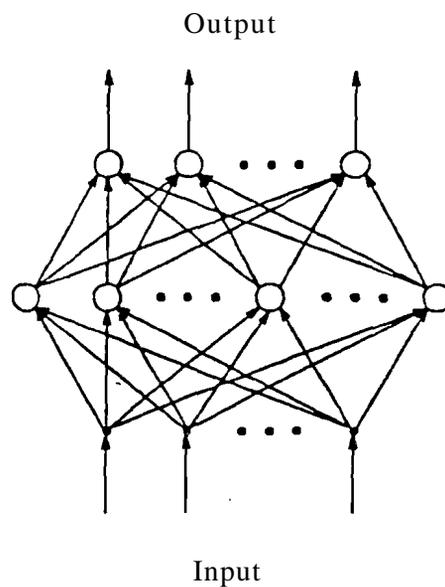


Fig.4.2 A multilayer perceptron network with one hidden layer.

Connections within a layer or from higher to lower layers are not permitted. Each node in a layer is connected to all the nodes in the layer above it. Training consists of finding a set of weights for all the connections such that the desired output is generated for each input. When the MLP is used as a classifier, all output nodes are set to 0 except for the node that corresponds to the class to which the input pattern belongs. The desired output for that node is 1.

In our study, we have used a feedforward network with 6 nodes in the input layer corresponding to the moment invariant features and 20 nodes in the output layer corresponding to the objects. It has one hidden layer in between the input and the output layers. For classification, the network is trained to the set of noise free normalised patterns.

4.4.3 Error back propagation algorithm

An effective iterative gradient-descent procedure has been developed for training MLPs. Learning proper weights using this algorithm is done as follows:

For each pattern in the training set, compute the error between the desired and the actual outputs and feed back this error signal level by level to the inputs changing each weight in proportion to its responsibility for the output error.

The detailed algorithm is as follows:

- Step 1:** Initialise all w_{ij} 's to small random values with w_{ij} being the connection weight between unit j and unit i in the layer below.
- Step 2:** Present an input pattern from class m and specify the desired output. The desired output is 0 for all the output nodes except the m th node which is 1.
- Step 3:** Calculate actual outputs of all the nodes using the present value of w_{ij} . The output of node j , denoted by y_j , is a nonlinear function of its total input:

$$y_j = \frac{1}{1 + \exp(-\sum_i y_i w_{ij})}$$

This particular nonlinear function is called a sigmoidal function.

Step 4: Find an error term, δ_j , for all the nodes. If d_j and y_j stand for, respectively, the desired and actual value of a node, then for an output node,

$$\delta_j = (d_j - y_j) y_j (1 - y_j)$$

and for a hidden layer node,

$$\delta_j = y_j (1 - y_j) \sum_k \delta_k w_{jk}$$

where k indexes all nodes in the layer above node j .

Step 5: Adjust weights by

$$w_{ij}(n+1) = w_{ij}(n) + \alpha \delta_j y_i + \gamma (w_{ij}(n) - w_{ij}(n-1))$$

where the **indices** $(n+1)$, (n) and $(n-1)$ represent next, **present** and previous respectively. The parameter α is a learning rate similar to step size in gradient search algorithms, and γ is a constant between 0 and 1 which determines the effect of past weight changes on the current **direction** of movement in the weight space. This provides a kind of momentum that effectively filters out high frequency variations of the error surface.

Step 6: Present another input and go back to Step 2. All the training inputs are presented cyclically until weights converge.

This algorithm is an iterative gradient descent procedure in the weight space which minimizes the total error between the desired and actual outputs of all the nodes in the **system**.

4.5 EXPERIMENTAL STUDIES

4.5.1 Experiments in Classification

Eight different images from each of the twenty olympic games symbols are generated, consisting of varying scales, orientations and translations of each image. Some of these images (six for each symbol) are shown in Fig.4.3.

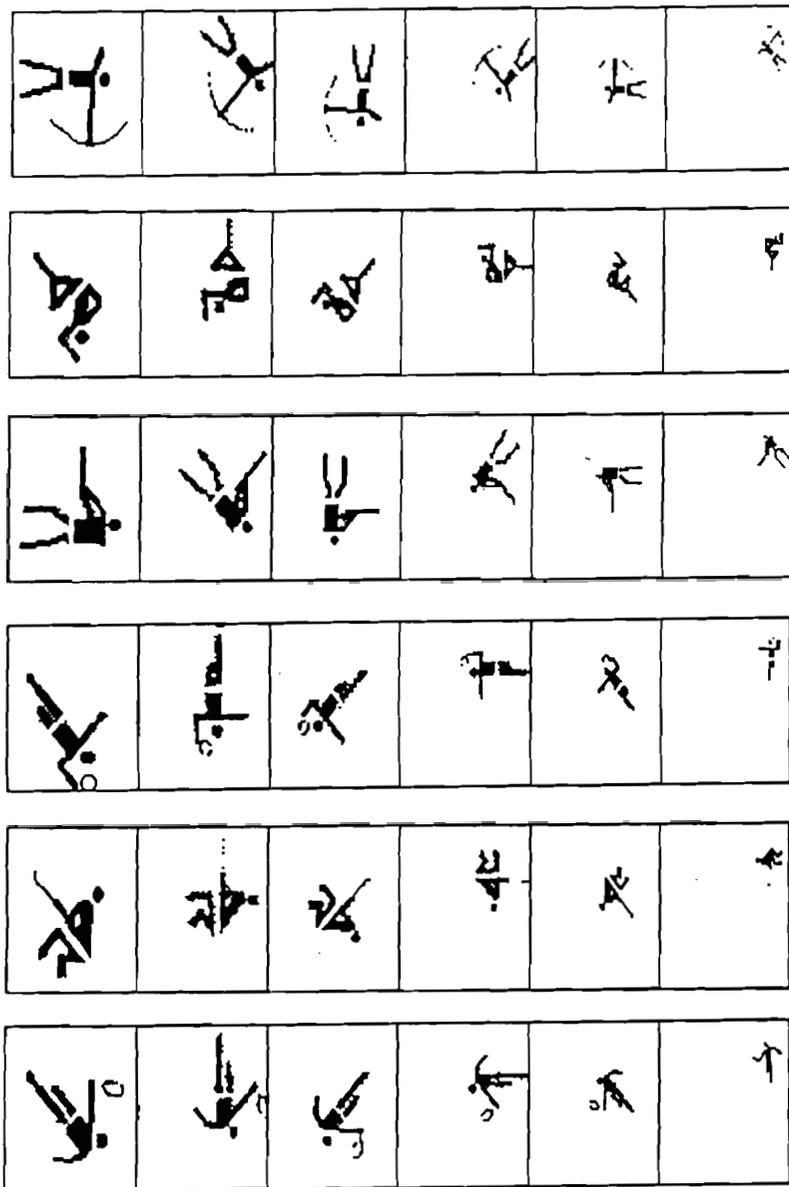


Fig.4.3 Some rotated, scaled and translated images of olympic game symbols used to study transformationinvariant recognition of objects.

Two images per symbol were used for training and remaining six for testing.

Fig.4.4 shows classification results for these images as they are scaled down in size besides other transformations. We have found that since reduction in size causes loss of details, classification accuracies of 100% are obtained upto a scale reduction which causes a 1:1/3 reduction in the length of the image (that is, for an original image size of 128x128 points, 1:1/3 reduction results in an image of approximately 40x40 points). Beyond this, further reduction causes wrong classification in many cases. Hence, for the set of olympic games symbols this appears to be the level upto which the approach can be reliably used.

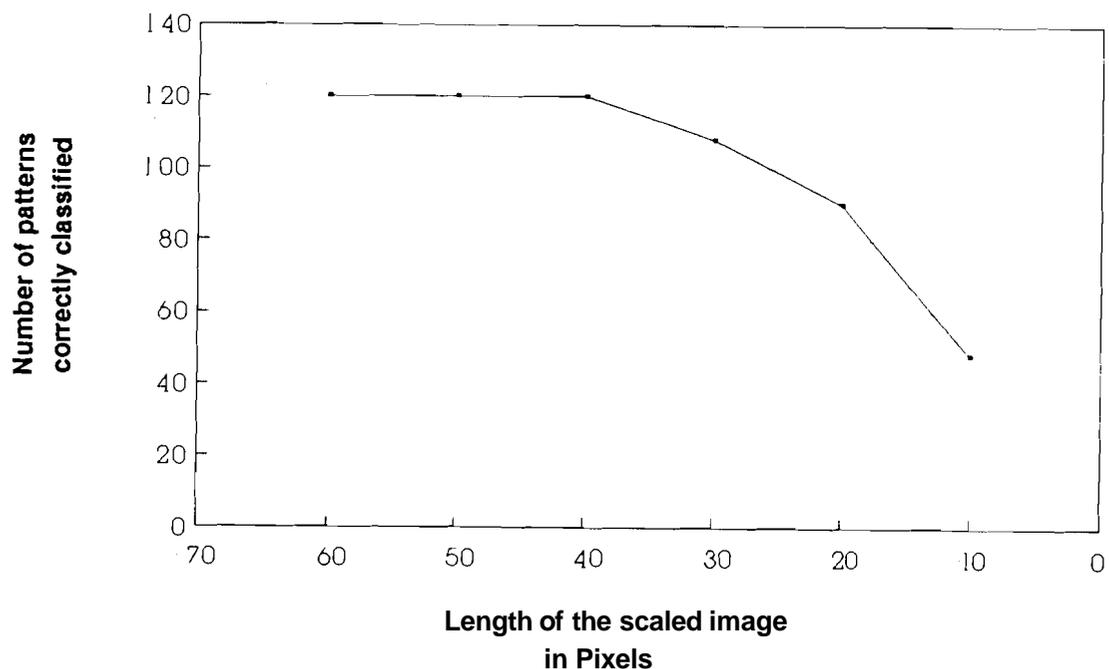


Fig.4.4 Transformation invariant recognition performance in the case of olympic symbols. Graph shows the number (out of a set of 120 test patterns) of objects (maximum size 128x128) correctly classified as the size of the image is reduced.

To see the effect of details in the image on recognition performance, we tested the approach in classification of characters of the alphabet. Fig.4.5 shows the ten characters used for this study. These images consist of simple features so that they can be scaled down over a wide range without severe distortion. Eight different images from each of these characters were generated by scaling, rotating and translating it. Some of these are shown in Fig.4.6. Two of these were used for training and the remaining six for testing. Classification accuracies of 100% were found for all test data, that is, upto a scale reduction which causes a $1:1/12$ reduction in length of the image (that is, reduced image of size 10×10 points). Thus the moment feature approach works better in case of objects with simpler shapes.

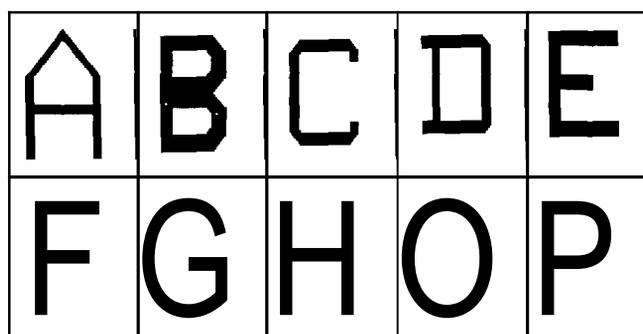


Fig.4.5 Images (128x128 points) of ten characters of the alphabet used in the study of transformation invariant object recognition.

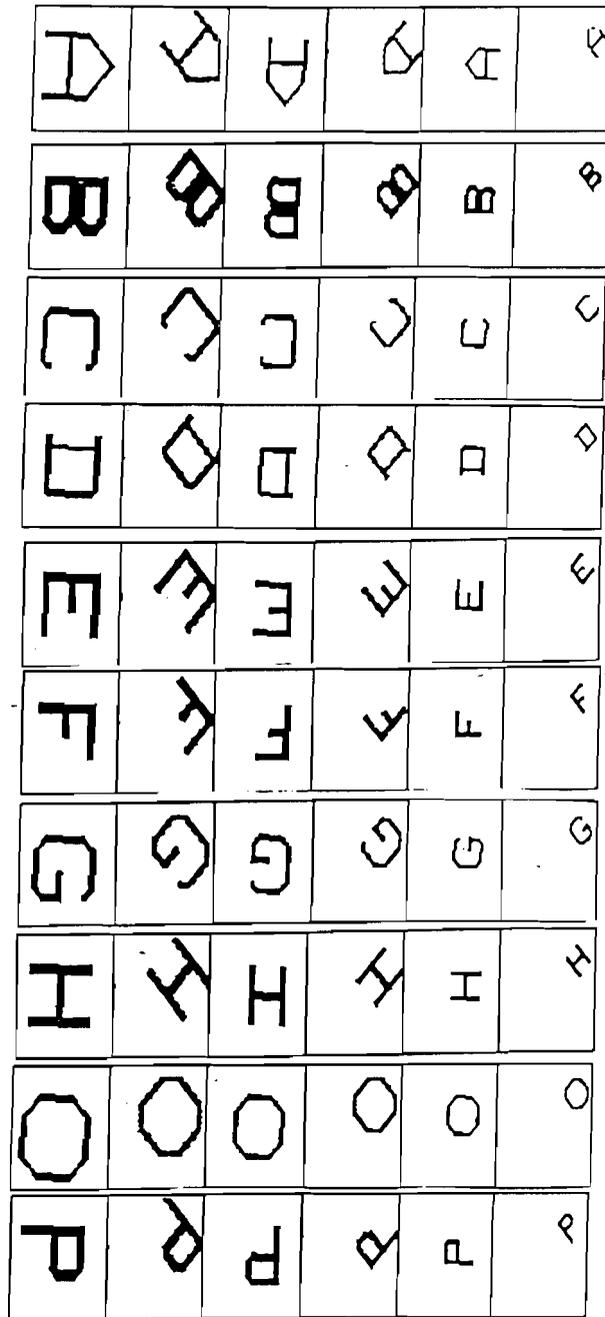


Fig.4.6 Some rotated, scaled and translated images of the characters generated to test transformation invariant recognition.

4.5.2 Experiments on the feature space

Further experiments were conducted using the alphabet image set. Robustness of moment features with respect to errors in centroid estimation is examined. Since all the six features that we use for classification are central moments whose accuracy depends on the correct estimation of the centroid, error was added to the centroid position. Moment invariants extracted based on this wrong estimate were used for classification. The results are shown in Fig.4.7.

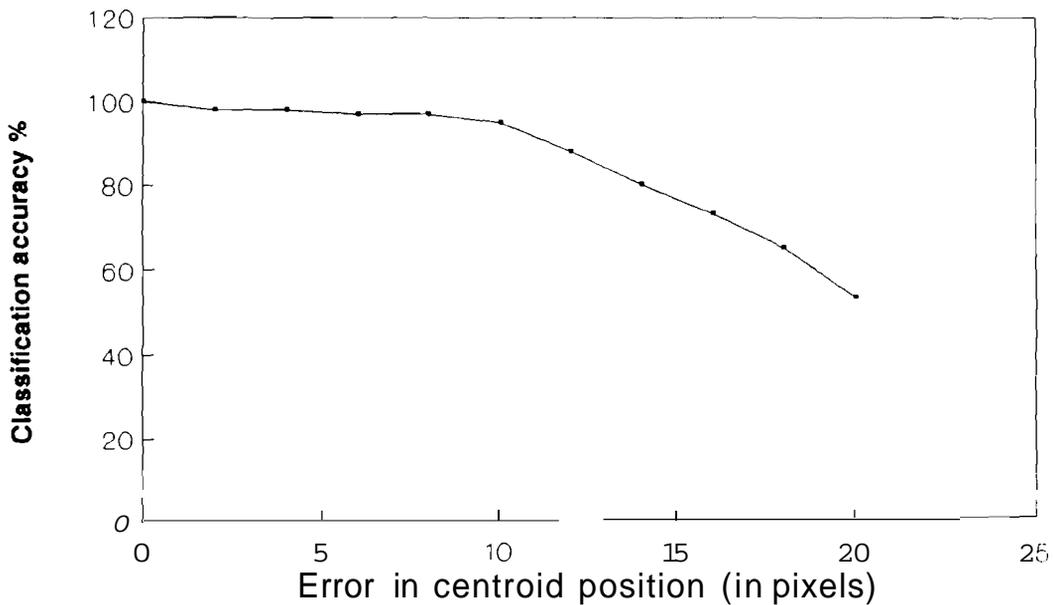


Fig.4.7 Recognition of images after introduction of error in centroid estimation. illustrating translation invariance of moment invariant features. The original image was of size 128x128 points.

Even when the centroid is displaced by 20 pixels, (original image defined on a 128x128 pixel grid), the classification accuracy falls to 96%. Hence moment invariant features are robust to effects of translation.

4.5.3 Experiments on the neural network classifier

Similarly, the generalisation capacity and fault-tolerance of the neural network

classifier were also tested.

Effect of number of training samples: To test the capacity of the network to generalise, training was done in two different ways. In the first case, half of the generated images (**i.e.**, four) for each class were used for training and the other half for testing. In the second case, only two images for each character were used for training and the rest (six) for testing. Performance changed very little when the number of training samples per class was reduced from 4 to 2.

Effect of increasing number of hidden layer nodes: The number of nodes in the hidden layer was varied and the classification accuracy was studied. We observed that around 8 hidden nodes, the peak classification accuracy of 100% is achieved. Increasing the number of hidden nodes beyond this does not have any effect on the performance.

Fault tolerance: In a network, processing is distributed among many nodes. Hence, even if some nodes fail to function properly, the effect on the overall performance of the system will not be appreciable. To test this capability of fault tolerance, we experimented by turning off several hidden layer nodes and observing the resulting effect on the system performance. The results are shown in Fig.4.8. The system exhibits a good degree of fault tolerance. Even when 25% of the hidden nodes are damaged, performance does not degrade significantly.

4.6 DISCUSSION

The results point out that if normalisation is done separately, then the tasks of feature extraction and classification are simplified. However, the approach has its limitations too:

- (i) This technique is not easily generalised to provide invariance against other nonlinear transformations of the pattern. For example, such an approach may not be suitable for hand printed character recognition where distortions are not necessarily linear transformations.
- (ii) In situations where the scene consists of multiple objects, the system must be capable of paying attention to individual objects in a scene for

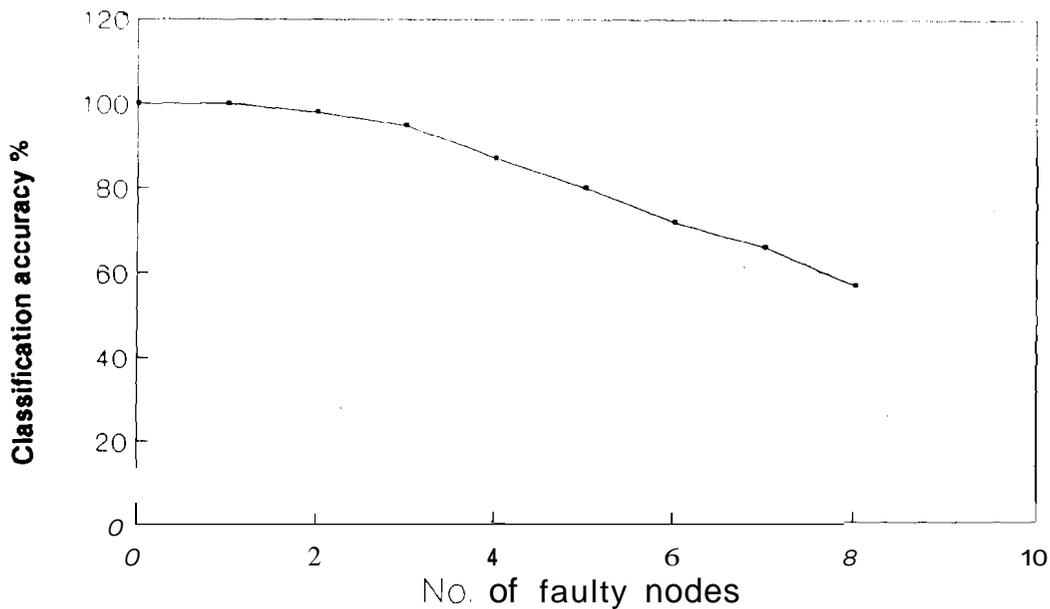


Fig.4.8 Degradation in system performance as a function of faulty hidden layer nodes. The network used twenty hidden layer nodes during training. Classification accuracy falls gradually with increasing number of faulty nodes, thus exhibiting a graceful degradation of performance. The study was conducted for the alphabet character set.

each of which the invariance of perception must separately be valid. In such cases, preprocessing of complete scenes with subsequent application of classification cannot constitute the whole solution.

Though the approach has these limitations, it **seems** effective in the present context, where we expect no more than one object in the scene at any time, the object is assumed to be rigid and the problem of occlusion is not addressed.

4.7 SUMMARY

Major Results : We have shown that a two-stage approach, where the tasks of **transformational** invariance and learned classification are handled separately, is successful in **recognising** objects over a wide range of scales and orientations. The method is less successful if images having finer details are scaled down significantly. We have also shown that if an invariant feature space is available, a multilayer neural network classifier can learn object shapes without explicit description.

Limitations: In the presence of noise the two-stage approach is inadequate. Noisy

and transformed images are difficult to recognise using the same approach. This is because during the computation of moments, we do not distinguish between pixels of the object and noise. If the moment features are extracted directly from the noisy image the estimates are not accurate, hence they give incorrect classification. There is thus a need for a preprocessing stage before features can be extracted. We address this case in the following chapter.

5

TRANSFORMATION INVARIANT RECOGNITION OF OBJECTS FROM DEGRADED IMAGES

In this chapter we address the issues of sparsity and noise together with transformational variability. We first discuss the need for preprocessing in Section 5.1. In Section 5.2 we describe the preprocessing approach, the aim of which is to reduce the noise and extract the object from the poorly resolved and noisy image. We propose a neural network to accomplish this task in Section 5.3. In Section 5.4 we describe experiments for recognition of objects from noisy reconstructed images. In Section 5.5 we discuss the results.

5.1 INTRODUCTION: NEED FOR PREPROCESSING

In the previous chapter we have studied transformation invariant recognition of objects in a simplified situation, where the issues of noise, poor resolution and other degradations did not exist. We have shown that in such a case, moment invariant features can be used to describe scaled, rotated and translated images and used in classification. However, in practice noise and imaging degradation exist. In this chapter we investigate to what level of degradation the recognition can be performed from noisy and degraded images.

Fig.5.1 illustrates the situation we are trying to address. Fig 5.1(a) shows the original image of an olympic games symbol. The transformed image obtained when the object moves is shown in Fig. 5.1(b). In the previous chapter we have shown that such

images can be recognised successfully. However, the noisy image that would be obtained by reconstruction from sparse data is shown in Fig.5.1(c). Such noisy and transformed images are difficult to recognise using the same approach, because during the computation of moments, we do not distinguish between pixels of the object and noise. If the moment features are extracted directly from the noisy image, the estimates give rise to incorrect classification. Hence there is a need for a preprocessing stage to reduce the noise and extract object pixels from the degraded image.

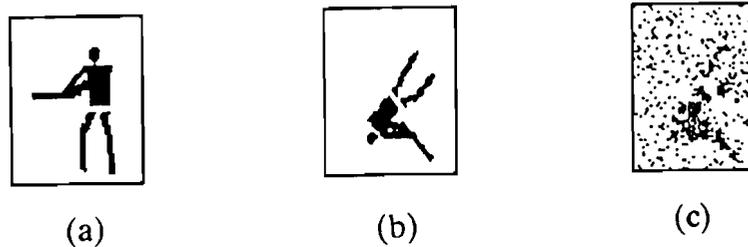


Fig.5.1 (a) Original image of an dypmic game symbol, (b) Transformed image obtained when the object moves and (c) Corresponding noisy and transformed image obtained by reconstruction from sparse data from a 32x32 sensor array.

We approach **the** problem of recognition of objects from degraded images in three stages, **consisting** of preprocessing, feature extraction and classification. This has to be compared and contrasted with the approach adopted in Chapter 3. We had argued that in the **case** of degraded images, feature extraction is unreliable and hence proposed a correlation matching approach. We have shown that if the task is only **classification**, then **this** approach can be used with severely degraded images. However, with **transformation** the task of recognition involves more than just classification. It involves **description and interpretation** at least in a primitive sense. Hence, we adopt a feature-classification approach for transformation invariant recognition of objects from noisy images.

However, it should be noted that such an approach is limited because feature extraction fails with increasing image degradation. Thus, while we are attempting to address increasing complexity in terms of transformational variability, we cannot expect comparable performance for equal amount of image degradation.

5.2 PREPROCESSING FOR NOISE SUPPRESSION AND OBJECT EXTRACTION : OUR APPROACH

Suppressing noise and segmenting an image into object in the foreground and noise in the background is a nontrivial task. In general, this requires physical and semantic knowledge about generic class of objects and even the specific object [GROSSBERG 89][LEVINE 87].

However, if the picture can be modelled as a single two-dimensional object superimposed on a uniform background, general purpose models may be useful. These include models for general classes of local features such as blobs, edges etc, as well as models that describe how such features can be grouped into **aggregates**[ROSENFELD 88] [SHER 91]. Traditional segmentation algorithms are two-stage sequential processes: local features are detected in the first stage, and they are grouped in the second stage.

In the presence of noise and data sparsity, such a strictly sequential process, where labels are first determined and processed later, may not work. This is because when data is uncertain, labels are ambiguous, **e.g.**, each pixel in a binary image may be interpreted as an image pixel or as a noise pixel. Hence, there is a need for an interactive process where locally ambiguous interpretations compete to achieve a globally **unambiguous** interpretation. A neural architecture is ideally suited for such a task.

We use a **multiscale** processing **approach**[ROSENFELD 71,84]. In the case of a compact rigid object with homogenous surface, measured surface properties remain same over a range of scales and over neighbourhoods, whereas noise is specific to one scale and location. In other words, surface features which are stable when the scale is varied can be considered as *features* of the object and those that disappear abruptly can be labelled as noise. This generalisation holds good for compact objects.

In the following section we propose an analog neural network in which measurements and hypothesis from differently sized neighbourhoods are integrated through cooperative and competitive interactions to perform noise suppression and object extraction.

5.3 A NEURAL NETWORK FOR PREPROCESSING

The proposed network consists of three stages. In the first stage surface patches are detected at three different scales by measuring local image contrast. A surface patch represents a region of the object. Due to presence of noise and the fact that the detector windows overlap, the detector outputs are ambiguous. Hence in stage two, by competitive interaction between adjacent detectors of same scale, ambiguity is reduced. In stage three the surface patches at different scales interact to obtain a combined output.

STAGE 1 : UNORIENTED CONTRAST DETECTION

Unoriented contrast detectors exist corresponding to every point in the image. Each contrast detector has an on-center off-surround receptive field as shown in Fig.5.2 and is sensitive to the amount and spatial scale of image contrast at a given image location. Thus each detector hypothesises a surface patch of its scale present in the image.

Let $I(\mathbf{x},\mathbf{y})$ denote the value of the input image at position (\mathbf{x},\mathbf{y}) in the lattice. The total excitatory input to the detector, $E_s(\mathbf{x},\mathbf{y})$, is obtained by integrating the total activation in the inner (on-center) receptive field:

$$E_s(\mathbf{x},\mathbf{y}) = \frac{\iint_{inner} I(\mathbf{x},\mathbf{y}) \, dx \, dy}{\iint_{inner} dx \, dy}$$

where s is the index of the size of the receptive field.

Similarly, The total inhibitory input, $F_s(\mathbf{x},\mathbf{y})$, is obtained by integrating the total activation in the outer (off-center) receptive field:

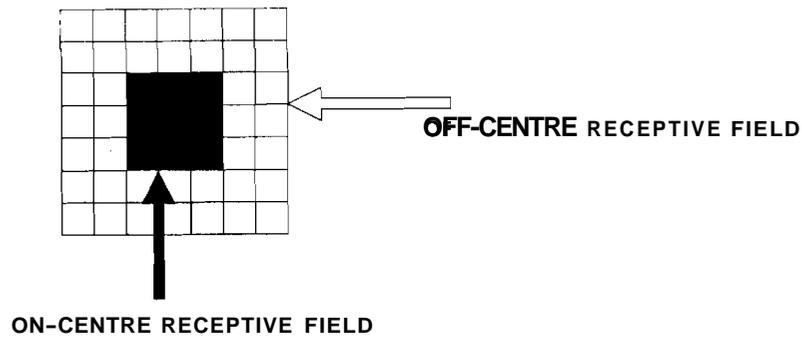


Fig.5.2 The kernel of the contrast detector used in the preprocessing stage, showing on-center and off-center receptive fields.

$$F_s(x,y) = \frac{\iint_{outer} I(x,y) dx dy}{\iint_{outer} dx dy}$$

The output of the contrast detector is defined as

$$C_s(x,y) = \max[E_s(x,y) - \alpha_s F_s(x,y) - \beta_s, 0],$$

where α_s is a contrast parameter such that $\alpha_s > 1$, β_s is a threshold parameter such that $0 < \beta < 1$ and the *mar* operator ensures that the output signal is nonnegative.

Large scale filters suppress noisy pixel distributions effectively, but localise the image patches poorly due to their broader spatial sampling. Small scale filters on the other hand, are sensitive to noise but are more reliable in localisation. Hence each filter in itself is insufficient to suppress noise and extract localised image patches. We need multiple scale interactions-These are described in stage 3.

STAGE 2: SPATIAL COMPETITION WITHIN EACH SCALE

The aim of this competitive stage, realised by an on-center off-surround network, is to reduce ambiguity in hypothesis among spatial neighbours and select more probable hypothesis. Each detector output excites the cell activity at the next layer

which represents the same position and scale, while inhibiting cell activities at the neighbouring locations as shown in Fig.5.3. **As** a result, cells that have high activity suppress activities of nearby cells which have lower activity due to noise. Under equilibrium state, the cells which initially had higher activity saturate in a **winner-take-all** fashion and those with low initial activity are cut-off.

The output $D_s(x,y)$ of the cell at position (x,y) and scales is given by

$$D_s(x,y) = \frac{C_s(x,y)}{1 + \gamma_s \sum_x \sum_y C_s(x,y)G_s(x,y)}$$

where $C_s(x,y)$ is the activity of the detector at the input, $G_s(x,y)$ is the competition kernel and γ_s is a parameter for controlling the effect of competition.

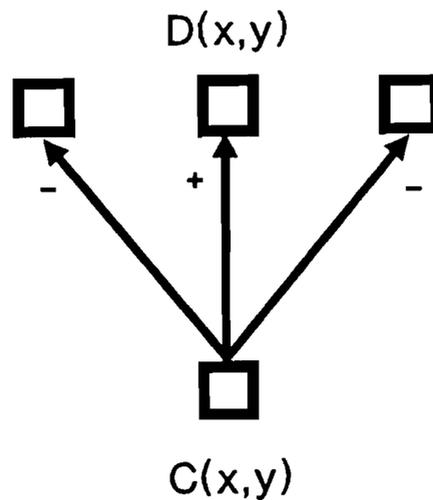


Fig.5.3 Spatial competition between detectors of the same scale: Each detector output $C(x,y)$ excites the cell activity $D(x,y)$ at the next layer which represents same position and scale and inhibits cell activities at the neighbouring locations.

STAGE 3 : MULTIPLE SCALE INTERACTION

The responses of the various detectors are combined so as to retain evidence that is available at multiple scales and remove those unsupported across scales. This is done through cooperative interactions.

$$M(x,y) = D_1(x,y) [\Sigma D_2(x,y) U(x,y)] [\Sigma D_3(x,y) U(x,y)]$$

where $D_1(x,y)$, $D_2(x,y)$ and $D_3(x,y)$ are the responses of detectors at three different scales in the present implementation, $U(x,y)$ is an unoriented excitatory kernel and $M(x,y)$ is the combined output. The purpose of the kernel is to make the effect of the larger scale detectors more diffuse spatially owing to their broader receptive fields. The noise-suppressed output $M(x,y)$ is analog, ranging from 0 to 1, and is thresholded to yield a binary image.

5.4 EXPERIMENTAL STUDIES IN PREPROCESSING AND CLASSIFICATION

5.4.1 Dataset

A subset of **olympic** games symbols consisting of 10 symbols shown in Fig.S.4 was chosen for this study. Noisy, incomplete and transformed images of these were obtained by reconstruction from simulated sparse data. We considered **32x32** array and 16x16 array data for preprocessing and classification studies.

Besides olympic games symbols, images of alphabet characters were also used to test the classification performance in the case of objects with simpler shapes.

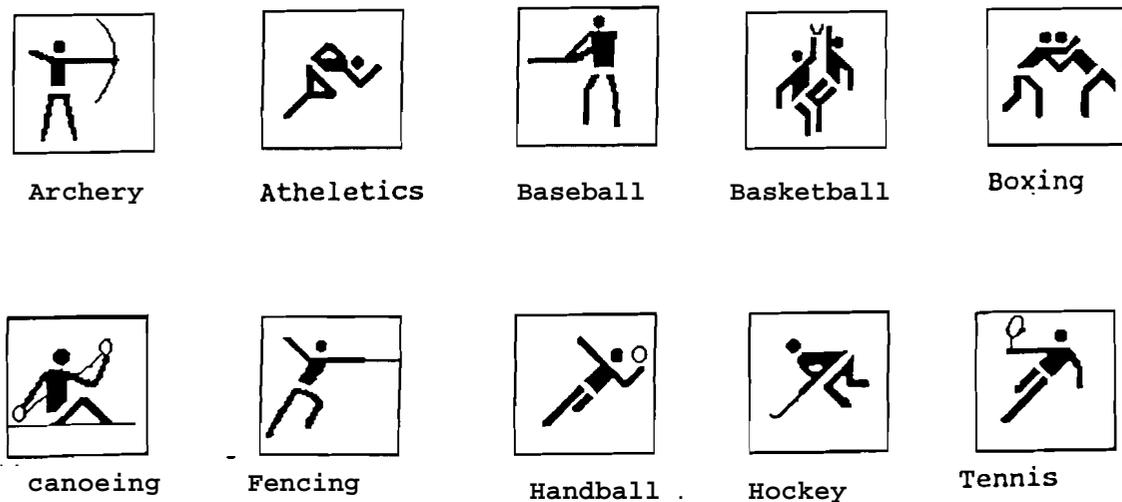


Fig.5.4 A subset of dlympic game symbols consisting of ten symbols used to study transformation invariant object recognition from noisy images

5.4.2 Preprocessing Performance

Fig 5.5 shows some images reconstructed from data collected by a 32x32 array along with the result of preprocessing. Fig.5.5(a) shows the transformed images used as objects in imaging simulation. Corresponding reconstructed images are shown in Fig.5.5(b) and preprocessed outputs are shown in Fig.5.5(c).

Even to the human observer, the preprocessed images in Fig.5.5(c) are much more clearer than those in Fig.5.5(b). This is because many unnecessary details have been removed by preprocessing, reducing the strain on the observer who can now concentrate his attention on the discriminating features. It is still difficult for us to recognise the objects from the preprocessed images due to poor resolution and missing parts. However, given the preprocessed image along with the list of the original symbols, the task of the human observer is simplified. This brings out the design

philosophy adopted in this thesis. Even though it may not be possible to rely on the system for stand-alone recognition performance, it should still be possible to use the output as an aid to human decision making.

Fig.5.6 shows the results when a sparser 16x16 sensor array is used for data collection. We observe that preprocessing results are not as good as for the case of 32x32 array data. In the preprocessed images, many parts are missing and there also some spurious parts. Still, it is interesting to see that preprocessing is effective in reducing noise from such degraded images, at least when the images are not scaled down significantly.

5.4.3 Classification performance

Transformation invariant features were extracted from the preprocessed image and are given as input to the neural network classifier for recognition. The results are as follows:

(i) *32x32 sensor array data for the case of olympic games symbols:*

Fig.5.7 shows the results. We observe that 100% recognition accuracy is obtained upto a scale reduction which causes 1:1/2 reduction in the length of the image (that is, when the reduced image is of size 50x50 pixels). This is to be compared with the recognition performance in the case of noise free images (See Fig.4.4 in Chapter 4) where 100% accuracies were obtained upto a scale reduction which causes 1:1/3 reduction in the length of the image. Hence with increased image degradation, the amount of distortion that can be tolerated is reduced.

(ii) *16x16 sensor array data in the case of olympic games symbols:*

Fig.5.8 shows the results. We observe that for a scale reduction which causes a 1:1/2 reduction in the length of the image, the classification accuracy is only about 65%. While this classification performance is quite impressive as the corresponding images look extremely poor in quality, this percentage is too low to be reliable in a practical system. Moreover, this falls very rapidly with further scale reduction such that for a reduction which causes a 1:1/4 recognition accuracy is nearly zero.

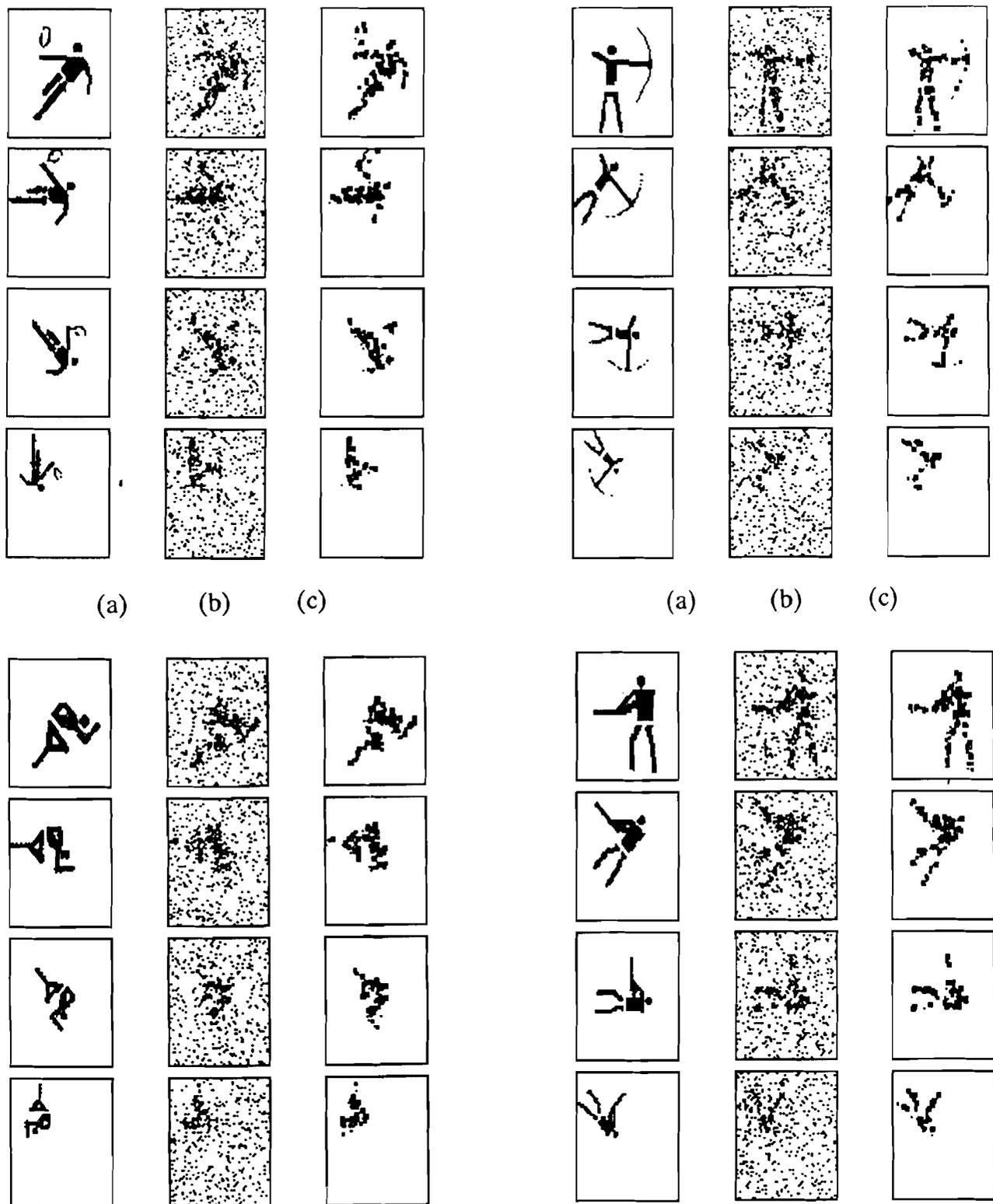


Fig.5.5 (a) Transformed images of some dlympic game symbols, tennis, archer, atheletics and baseball. (b) Corresponding images obtained by reconstruction from data collected by a sparse 32x32 sensor array at two different frequencies. (c) Images in (b) after noise suppression.

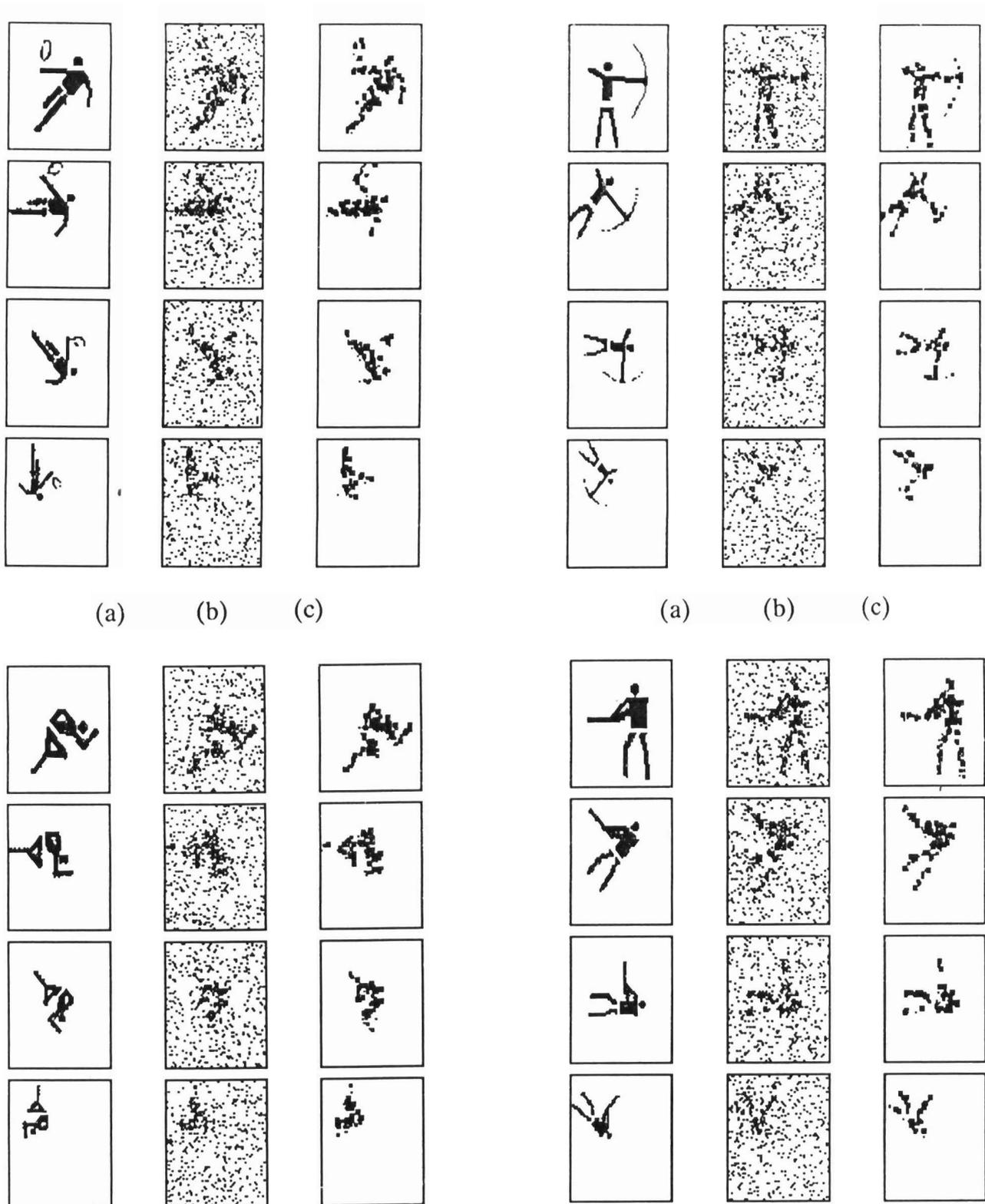


Fig.5.5 (a) Transformed images of some olympic game symbols, tennis, archer, athletics and baseball. (b) Corresponding images obtained by reconstruction from data collected by a sparse 32x32 sensor array at two different frequencies. (c) Images in (b) after noise suppression.

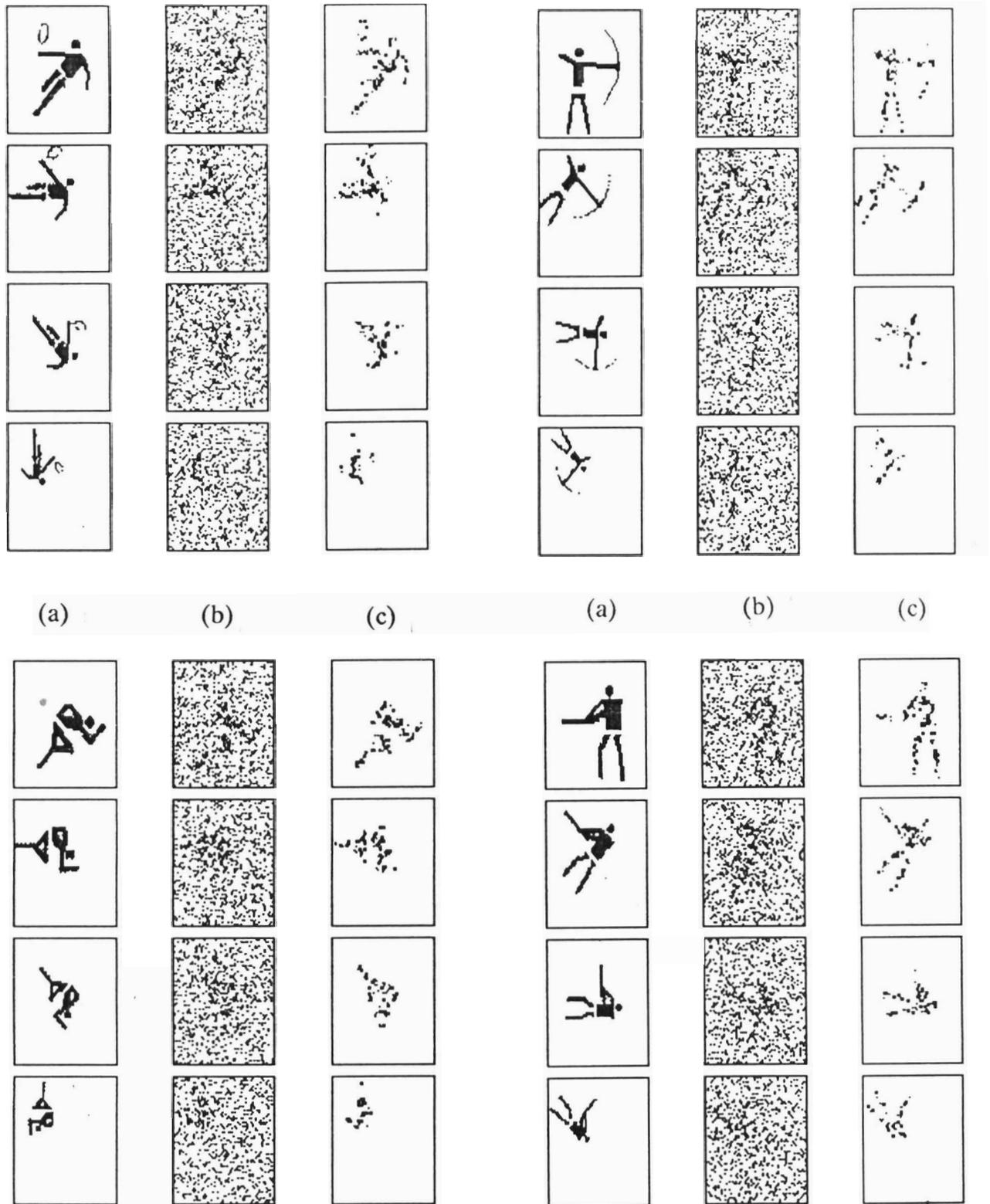


Fig.5.6 (a) Transformed images of some olympic game symbols, tennis, archer, athlete and baseball. (b) Corresponding noisy images obtained by reconstruction from data collected by a sparse 16x16 sensor array at two different frequencies. (c) Images in (b) after noise suppression.

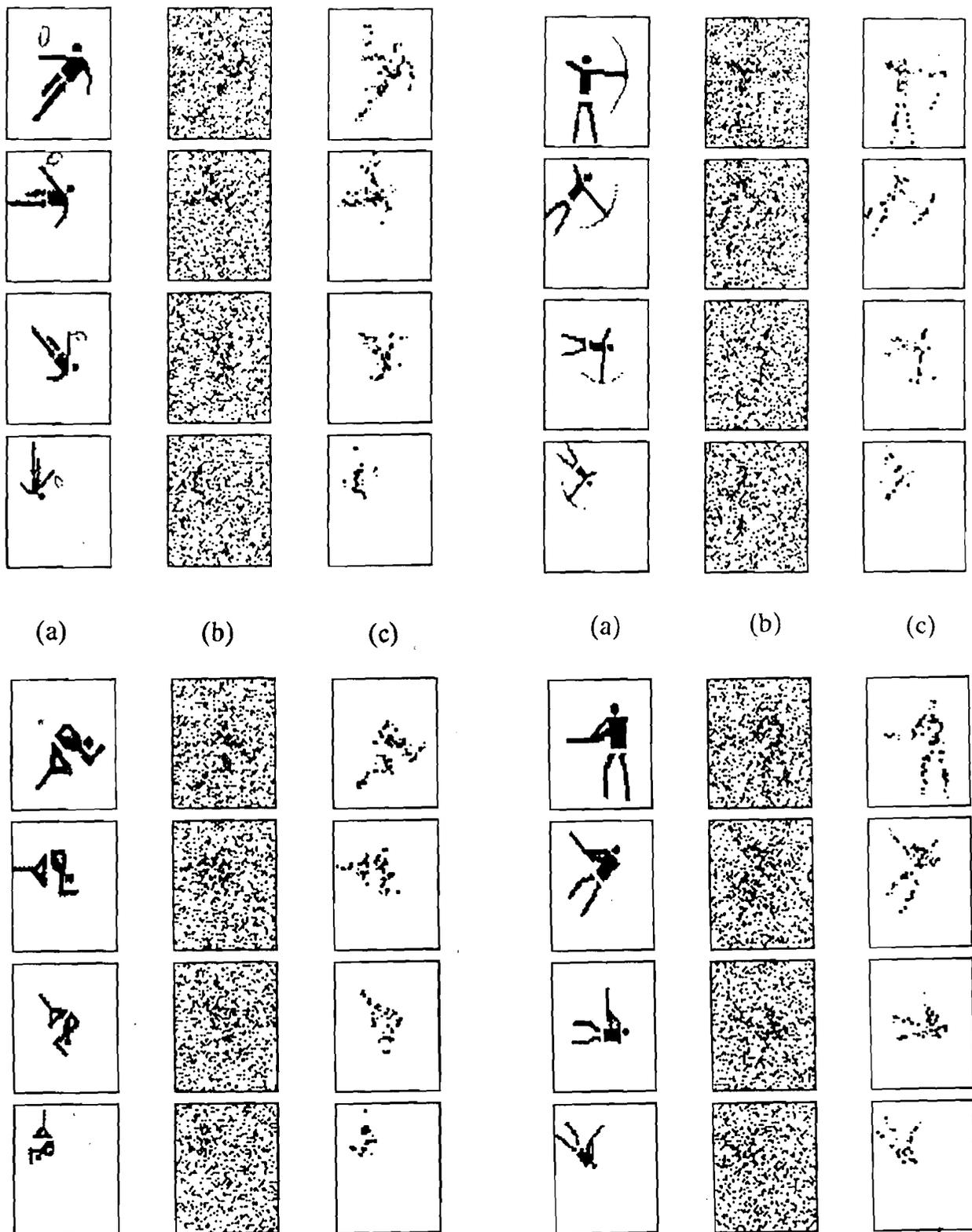


Fig.5.6 (a) Transformed images of some olympic game symbols, tennis, archer, athelitics and baseball. (b) Corresponding noisy Images obtained by reconstruction from data collected by a sparse 16x16 sensor array at two different frequencies. (c) images in (b) after noise suppression.

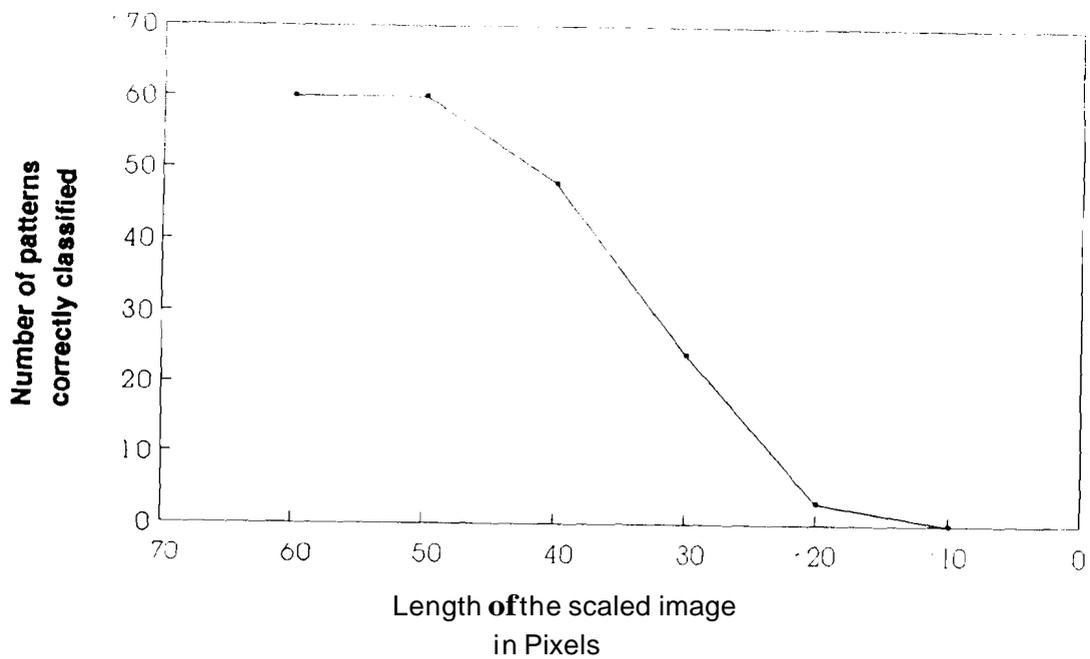


Fig.5.7 Transformation invariant recognition of olympic game symbols from images obtained by reconstruction from data collected by a 32x32 array at two frequencies. Graph shows the number (out of a set of 60 test patterns) of objects (maximum size 128x128) correctly classified as the size of the image is reduced.

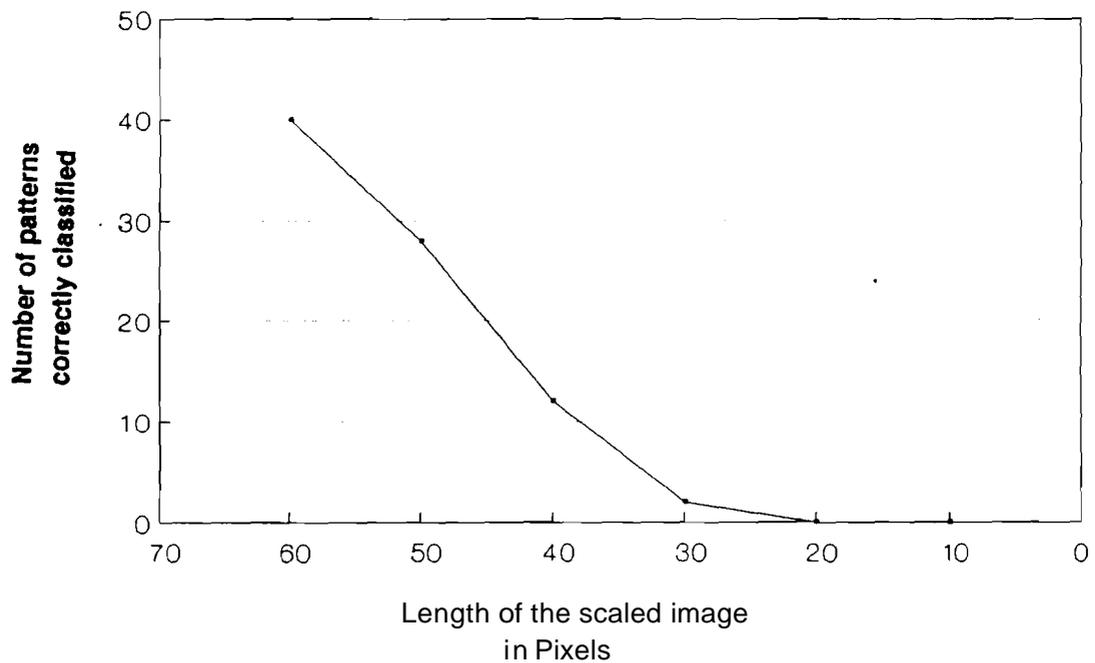


Fig.5.8 Transformation invariant recognition of olympic game symbols from Images obtained by reconstruction from data collected by a 16x16 array at two frequencies. Graph shows the number (out of a set of 60 test patterns) of objects (maximum size 128x128) correctly classified as the size of the image is reduced.

(iii) *32x32 sensor array data for the case of alphabet symbols:*

To test the recognition performance with a set of objects with simpler shapes, the experiment was performed on the alphabet data. Results are shown in Fig.5.9. 100% classification accuracy was obtained **upto** a scale reduction which causes a 1:1/3 reduction in length of the image (when size of the reduced image is 40x40 points), which is better than for olympic games symbols.

(iv) *16x16 sensor array data for the case of alphabet symbols:*

Results are shown in Fig.5.10. Near 100% classification accuracy was obtained only **upto** a scale reduction which causes a 1:1/2 reduction in length of the image (when size of the reduced image is around 60x60 points), which is better than for **olympic** games symbols 16x16 array data case, but worse than the alphabet 32x32 array data case.

5.5 DISCUSSION

In Chapter 3, we had addressed only imaging degradations and not considered transformational variability. Hence the present attempt can be viewed as making the problem more complex by introducing transformational variability. Results have shown that *for same level of image degradation, with transformational variability recognition performance degrades*. For example, images reconstructed from 16x16 array data were classified correctly with **100%** accuracy when transformations were not present. With **transformations**, recognition performance gracefully degrades. Classification accuracy is only around 65% for a scale reduction which causes a 1:1/2 reduction in image length.

In Chapter 4, we had addressed only transformational variability and not considered imaging degradations. Hence the present attempt can be viewed as making the problem more complex by introducing degradations. Results show that *for the same level of transformational distortion, increasing degradation reduces the recognition performance*.

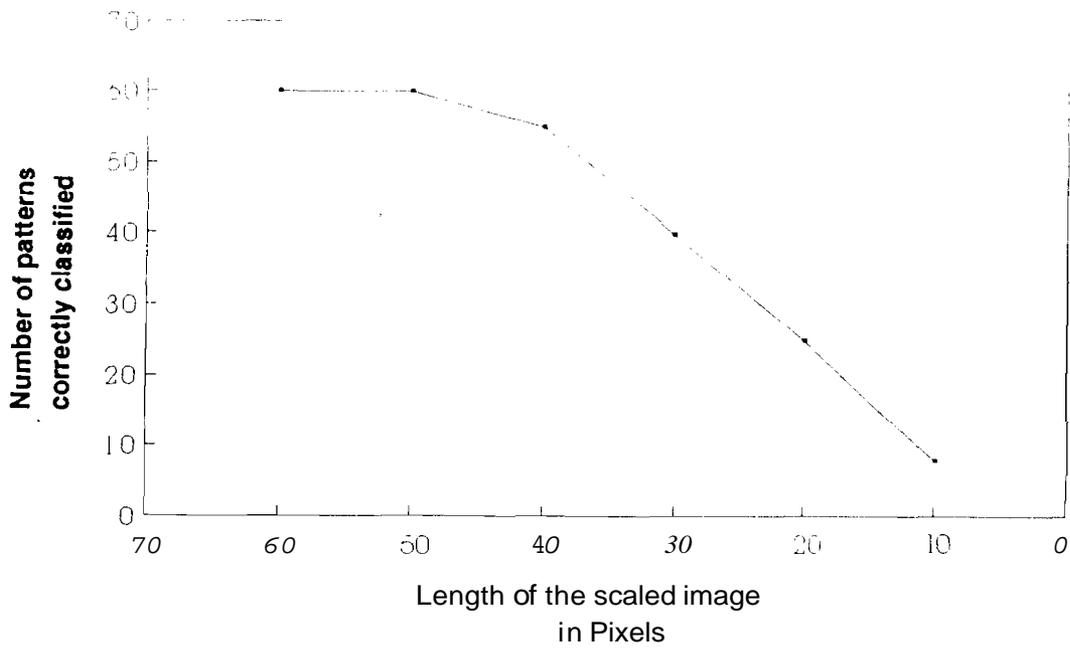


Fig.5.9 Transformation invariant recognition of alphabet characters from images obtained by reconstruction from data collected by a 32x32 array at two frequencies. Graph shows the number (out of a set of 60 test patterns) of objects (maximum size 128x128) correctly classified as the size of the image is reduced.

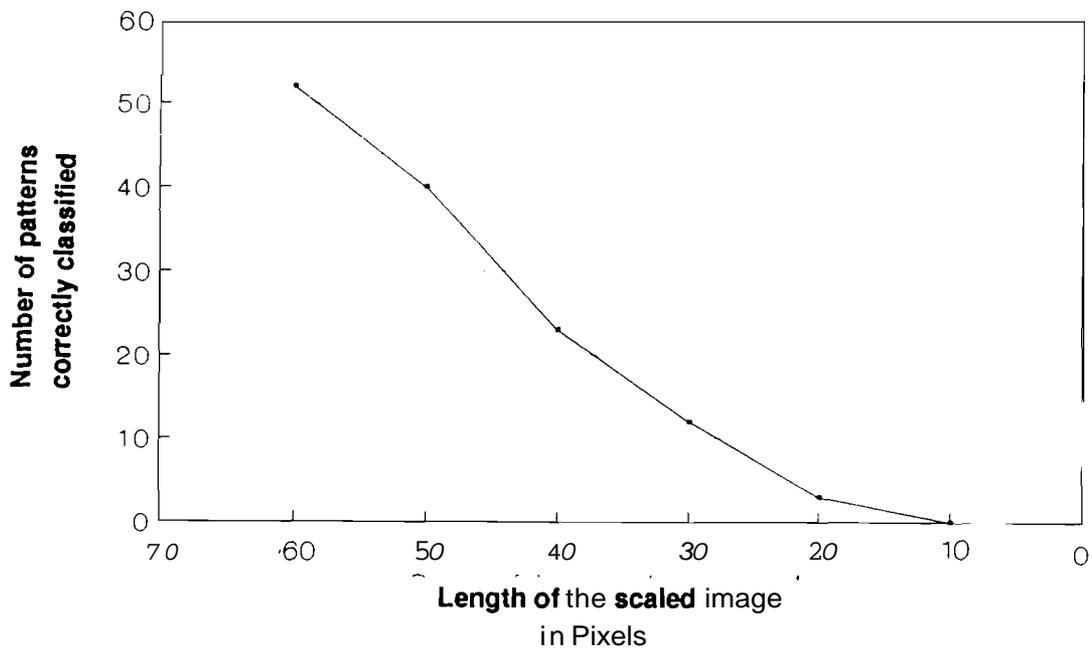


Fig.5.10 Transformation invariant recognition of alphabet characters from images obtained by reconstruction from data collected by a 16x16 array at two frequencies. Graph shows the number (out of a set of 60 test patterns) of objects (maximum size 128x128) correctly classified as the size of the image is reduced.

Increased complexity of the recognition task either due to transformational variability or due to imaging degradation causes the performance to degrade. This observation reflects the fact that transformation invariant recognition of objects from noisy images involves more than just classification - it requires description and interpretation. The approach can still be used for object recognition from noisy transformed images even though much smaller degradation only can be tolerated. For example, the present approach can be used effectively for transformation invariant recognition of **olympic** games symbols from images reconstructed from 32x32 array data upto a scale reduction corresponding to 1:1/2 reduction in length of the image.

The graceful degradation characteristic is desirable in practice due to the following reasons:

- (i) It reflects the reliability of a system
- (ii) Even when the system cannot give a reliable output decision, its intermediate results can be used as an aid to human decision making. For example, a human expert can improve his performance if he is given along with the reconstructed images, the preprocessed images also.

5.6 SUMMARY

In this chapter, we have addressed the task of recognising objects from degraded and transformed images. The need for a feature preprocessing stage was stressed. A neural network architecture was proposed for preprocessing. This network uses lateral cooperation and competition between nodes with receptive fields of various sizes at the same level to achieve noise suppression. We have demonstrated that this neural network preprocessing is very effective in overcoming the effects of noise and sparsity of data. Using this preprocessing, transformation invariant object recognition can be obtained from noisy images. The recognition performance degrades gracefully with increasing complexity - due to imaging degradation and due to increasing details of the object.

6

SUMMARY

6.1 SUMMARY OF THE THESIS

In this thesis, we have tried to explore the possibility of developing *artificial neural network models* that can be trained to identify objects from poorly resolved, noisy and transformed (scaled, rotated and translated) images, such as images reconstructed from sparse and noisy data. Our motivation was to use the adaptive processing capabilities of neural networks, both as classifiers and as context sensitive processors, to address issues of data uncertainty and variability due to noise, incompleteness and distortion.

Studies reported here were made by simulating noise and sparsity of data as obtained in a simplified model of a *sensor array imaging* situation. Noise and sparsity of data degrade the quality of the image at every pixel instead of affecting it in the form of local corruption of image information as in many image processing situations. Hence, in the case of images reconstructed from sparse and noisy data, (i) neighbourhood processing methods for noise cleaning are not applicable, (ii) feature extraction cannot be reliably performed, and (iii) model based methods for classification cannot be applied for the same reasons. We have shown in this thesis through simulation studies that neural network models can overcome some of these limitations by learning and context sensitive processing.

We have described the studies in object recognition for three different cases:

- (i) In the first case, we have considered the issues of noise in the received data and sparsity of the data. We have described studies on *object recognition from degraded images* using a simple trained neural network

which performs correlation matching.

- (ii) In the second case, we have considered the issue of *transformations* alone. We have described a feature space that is invariant to transformations, and a neural network classifier for recognition.
- (iii) In the third case, we have addressed the issues of noise and transformation together and studied *transformation invariant recognition of objects from degraded images*. For the first two cases existing neural network architectures were explored for recognition of objects. For the third case new methods were proposed for preprocessing and recognition.

6.2 MAJOR RESULTS OF THE THESIS

We have studied **object recognition from degraded images** as it occurs in situations such as sensor array imaging:

- (i) We have shown that if the set of expected objects are known, then it is possible to train a neural network for object recognition from **degraded** images if transformational variability is not present. Even though the present network performs a simple correlation matching, it classifies even images which are so degraded that we fail to perceive the discriminating features visually. Thus, this study demonstrates that in some situations, a trained neural network classifier can perform object recognition very well in the presence of degradation.
- (ii) If image degradations are not present, **transformation invariant recognition of objects** can be performed by a trained neural network based on invariant feature measures. We have also shown that this two-stage approach, where the tasks of transformational invariance and learned classification are handled separately, is successful in recognising objects over a wide range of scales and orientations. However, the method is less successful if the images have finer details and are severely scaled down.
- (iii) To enable **transformation invariant object recognition from degraded images**, we have proposed a neural network *preprocessing stage* to

overcome the effects of noise and sparsity. This network uses context sensitive lateral cooperation and competition between nodes with receptive fields of various sizes to achieve noise suppression.

We have demonstrated the **advantages of using neural networks** for this complex task of pattern recognition:

- (i) Unlike model based methods, which have to be tailored to specific types and even levels of degradation, neural network classifiers work with more than one type of degradation.
- (ii) Neural network classifiers can learn object shapes **without explicit description**.
- (iii) Neural network classifier exhibits desirable characteristics of **fault tolerance, graceful degradation** of performance with increasing distortion etc.

We have made a **systematic study** of the object recognition problem by identifying aspects of complexity, namely, types and levels of image degradation, transformational variability, nature of symbol set etc., and performing the study in stages of increasing complexity. This enables identifying the **limits upto** which the systems and approaches can be useful.

- (i) The systematic study has revealed a desirable characteristic of neural network models, namely, **graceful degradation** of performance with increasing complexity. That is, neural architectures can perform well **upto** a level beyond which the degradation is gradual.
- (ii) In situations where complexity is less, these neural network based methods can be used as standalone object recognition schemes. As the complexity of the recognition task increases due to increased detail in the image, and/or due to degradations in imaging, the output of the system can still be used as an **aid to human decision making**.

REFERENCES

- [BALLARD 82] D. H. Ballard and C. M. Brown, *Computer Vision*, Englewood Cliffs, NJ: Prentice-Hall, 1982.
- [BURT 88] P.J.Burt, "Smart sensing within a pyramid vision machine," *Proceedings of the IEEE*, vol.76, pp.1006-1015, August 1988.
- [BARNARD 91] E.Barnard and D.Casasent, "Invariance and neural nets," *IEEE Transactions on Neural Networks*, vol.2, pp.498-508, September 1991.
- [DUDANI n] S.A.Dudani, K.J.Breeding, and R.B.McGhee, "Aircraft identification by moment invariants," *IEEE Transactions on Computers*, vol. C-26, pp.39-45, January 1977.
- [FISCHLER 87] M.A.Fischler and O.Firschein, *Intelligence-The Eye, the Brain and the Computer*, Addison-Wesley Publishing Co., 1987.
- [FUKUSHIMA 83] K.Fukushima, "Neocognitron: A neural network for a mechanism of visual pattern recognition," *IEEE Transactions on Systems, Man and Cybernetics*, vol.13, pp.826-834, 1983.
- [GRACE 91] A. E. Grace and M. Spann, "A comparison between Fourier-Mellin descriptors and moment based features for invariant object recognition using neural networks," *Pattern Recognition Letters*, vol. 12, pp. 635-643, October 1991.
- [GROSSBERG 80] S.Grossberg, "How does a brain build a cognitive code?," *Psychological Review*, vol.87, pp.1-51,1980.
- [GROSSBERG 89] S.Grossberg, E.Mingolla and D.Todorovic, "A neural network architecture for preattentive vision," *IEEE Transactions on Biomedical Engineering* vol. 36, pp.65-84, January 1989.

- [GILES 88] C.L.Giles, R.D. Griffin, and T. Maxwell, "Encoding geometric invariances in higher order neural networks," In *Neural Information Processing Systems*, D.Z.Anderson, (Ed.), Denver, pp.301-309, 1988.
- [HU 62] M.Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. IT-8, pp.179-187, February 1962.
- [KHOTANZAD 90] A.Khotanzad and J.Lu, "Classification of invariant image representations using a neural network," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol.38, June 1990.
- [LIPPMANN 87] R.P.Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Magazine*, vol.2, pp. 4-22, April 1987.
- [LINSKER 88] R.Linsker, "Self-organisation in a perceptual network," *IEEE Computer*, pp.105-117, March 1988.
- [MAITRA 79] S.Maitra, "Moment invariants," *Proceedings of IEEE*, vol.67, pp.697-699, 1979.
- [PROKOP 92] R.J.Prokop and A.P.Reeves, "A survey of moment-based techniques for unoccluded object representation and recognition," *CVGIP: Graphical Models and Image Processing*, vol. 54, pp.438-460, September 1992.
- [RAVICHANDRAN 91] A.Ravichandran and B.Yegnanarayana, "A two-stage neural network for translation, rotation and size-invariant visual pattern recognition," *Proceedings of the ICASSP 91*, Toronto, vol.4, pp.2393-2396, May 1991.
- [ROSENFELD 71] A.Rosenfeld and M.Thurston, "Edge and curve detection for visual scene analysis," *IEEE Transactions on Computers*, C-20, pp.512-569, 1971.
- [ROSENFELD 84] A.Rosenfeld, *Multiresolution Image Processing and Analysis*, Springer-Verlag, Berlin, 1984.

- [ROSENFELD 88] A.Rosenfeld and A.Sher, "Detection and delineation of compact objects using intensity pyramids," *Pattern Recognition*, vol.21, pp.147-151, 1988.
- [RUMELHART 86A] D.E.Rumelhart and J.L.McLelland, Eds. *Parallel Distributed Processing* Cambridge, MA : MIT Press,1986.
- [RUMELHART 86B] D.E.Rumelhart, G.E.Hinton, and R.J.Williams, "Learning internal representations by error propagation," In *Parallel Distributed Processing*. Cambridge, M A : MIT Press, pp.318-36,1986.
- [SHER 91] A.Sher and A.Rosenfeld, "Pyramid cluster detection and delineation by consensus," *Pattern Recognition Letters*, vol.12, pp.477-482, August 1991.
- [SMITH 71] F.W.Smith and M.H.Wright, "Automatic ship photo interpretation by the method of moments," *IEEE Transactions on Computers*, vol. C-20, pp.1089-1095, September 1971.
- [TEAGUE 80] M.R.Teague, "Optical calculation of irradiance moments," *Applied Optics*, vol.19, pp.1353-1356, 1980.
- [TEH 88] C.H.Teh and R.T.Chin, "Image Analysis by the method of moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.10, pp.496-513, July 1988.
- [TSOTSOS 84] J.K.Tsotsos, "Knowledge and the visual process: Content, form and use," *Pattern Recognition*, vol.17, pp.13-27, January 1984.
- [WIDROW 88] B.Widrow and R.Winter, "Neural nets for adaptive filtering and adaptive pattern recognition," *IEEE Computer*, vol.21, pp.25-39, 1988.
- [YEGNANARAYANA 91] B.Yegnanarayana, R.Ramaseshan and A.Ravichandran, *Studies in Sensor Array Imaging*, Indian Institute of Technology, Madras, 1991.
- [YIN 81] B.H.Yin and H.Mack, "Target classification algorithms for video and FLIR imagery," *Proceedings of SPIE*, vol. 302, pp.134-140, 1981.

LIST OF FIGURES

- Fig.1.1 A simplified Sensor Array Imaging setup.
- Fig.1.2 (a) Some Olympic game symbols (128x128 points) used as planar objects in Sensor Array Imaging simulation studies (b) Corresponding images (128x128 points) reconstructed from data collected by a sparse (16x16 sensors) array.
- Fig.1.3 Overview of the research showing simplified situations where selected issues are considered.
- Fig.3.1 The Hamming network. X's represent the input nodes and Y's represent the output nodes.
- Fig.3.2** Olympic game symbols (20) in the knowledge base used for the studies in recognition of objects from **degraded** images.
- Fig.3.3 Recognition results for images reconstructed from data collected at two frequencies from a 16x16 array. The class decision of the network is given along with the activation value of the winning pattern. In this case, all of the twenty images were correctly identified.
- Fig.3.4** Recognition results for images reconstructed from data collected at two frequencies from an **8x8** array. The class decision of the network is given along with the activation value of the winning pattern.
- Number of images correctly identified = 13.
- Fig.3.5** Summary of recognition performance with different sparse arrays (64x64, 32x32, 16x16 and 8x8 sensors). Graph shows the number of patterns correctly identified out of twenty patterns in each case.
- Fig.3.6 Recognition results for images reconstructed from data collected at four frequencies from an 8x8 array. The class decision of the network is given along with the activation value of the winning pattern. Misclassification is underlined.
- Number of images correctly identified = 19.

Fig.3.7 Summary of recognition performance with different cases of multiple frequency data collection, for 8, 4, 2 and 1 frequencies used for data collection using an **8x8** sensor array. Graph shows number of patterns correctly identified out of twenty patterns in each case.

Fig.3.8 Recognition results for images reconstructed from noisy data (SNR = -3dB) collected at two frequencies from a **16x16** array. Noise was Gaussian and added to the received data. The class decision of the network is given along with the activation value of the winning pattern. Misclassification is underlined.

Number of images correctly identified = **18**.

Fig.3.9 Recognition results for images reconstructed from noisy data (SNR = -6dB) collected at two frequencies from a **16x16** array. Noise was Gaussian and added to the received data. The class decision of the network is given along with the activation value of the winning pattern. Misclassification is underlined.

Number of images correctly identified = **12**.

Fig.3.10 Summary of recognition performance with different levels of noise in the received data collected at two frequencies by a **16x16** array. Graph shows number of patterns correctly identified out of twenty patterns in each case.

Fig.3.11 Images reconstructed from sensor array data collected when the objects have moved **relative** to the array.

Fig.4.1 Several **olympic** game symbols represented in a two dimensional feature space formed by the first two moment invariant features, ϕ_1 and ϕ_2 . Values represent logarithm of the absolute values of the features, **normalised** to unity.

Fig.4.2 A multilayer perceptron network with one hidden layer.

Fig.4.3 Some rotated, scaled and translated images of olympic game symbols used to study transformation invariant recognition of objects.

Fig.4.4 Transformation invariant recognition performance in the case of olympic symbols. Graph shows the number (out of a set of **120** test patterns) of objects (maximum size **128x128**) correctly classified as the size of the image is reduced.

- Fig.4.5 Images (128x128 points) of ten characters of the alphabet used in the study of transformation invariant object recognition.
- Fig.4.6 Some rotated, scaled and translated images of the characters generated to test transformation invariant recognition.
- Fig.4.7 Recognition of images after introduction of error in centroid estimation, illustrating translation invariance of moment invariant features. The original image was of size 128x128 points.
- Fig.4.8 Degradation in system performance as a function of faulty hidden layer nodes. The network used twenty hidden layer nodes during training. Classification accuracy falls gradually with increasing number of faulty nodes, thus exhibiting a graceful degradation of performance. The study was conducted for the alphabet character set.
- Fig.5.1 (a) Original image of an olympic **game symbol**, (b) Transformed image obtained when the object moves and (c) **Corresponding** noisy and transformed image obtained by reconstruction from **sparse** data from a 32x32 sensor array.
- Fig.5.2 The kernel of the contrast detector used in the preprocessing stage, showing on-center and off-center receptive fields.
- Fig.5.3 Spatial competition between detectors of the same scale: Each detector output $C(x,y)$ excites the cell activity $D(x,y)$ at the next layer which represents same position and scale and inhibits **cell** activities at the **neighbouring** locations.
- Fig.5.4 A subset of **olympic game symbols** consisting of ten symbols used to study transformation invariant object **recognition** from noisy images
- Fig.5.5 (a) Transformed images of some dympic game symbols, tennis, archer, atheletics and baseball. (b) Corresponding images obtained by reconstruction from data **collected** by a **sparse** 32x32 sensor array at two different frequencies. (c) Images in (b) after noise suppression.

Fig.5.6 (a) Transformed images of some olympic game symbols, tennis, archer, athletics and baseball. (b) Corresponding images obtained by reconstruction from data collected by a sparse 16x16 sensor array at two different frequencies. (c) Images in (b) after noise suppression.

Fig.5.7 Transformation invariant recognition of olympic game symbols from images obtained by reconstruction from data collected by a 32x32 array at two frequencies. Graph shows the number (out of a set of 60 test patterns) of objects (maximum size 128x128) correctly classified as the size of the image is reduced.

Fig.5.8 Transformation invariant recognition of olympic game symbols from images obtained by reconstruction from data collected by a 16x16 array at two frequencies. Graph shows the number (out of a set of 60 test patterns) of objects (maximum size 128x128) correctly classified as the size of the image is reduced.

Fig.5.9 Transformation invariant recognition of alphabet characters from images obtained by reconstruction from data collected by a 32x32 array at two frequencies. Graph shows the number (out of a set of 60 test patterns) of objects (maximum size 128x128) correctly classified as the size of the image is reduced.

Fig.5.10 Transformation invariant recognition of alphabet characters from images obtained by reconstruction from data collected by a 16x16 array at two frequencies. Graph shows the number (out of a set of 60 test patterns) of objects (maximum size 128x128) correctly classified as the size of the image is reduced.

LIST OF PUBLICATIONS

Related Publications

1. B.Yegnanarayana, A.Ravichandran and R.Ramaseshan, "Object identification in images reconstructed from sensor array data," *Algorithms for Signal Reconstruction in Acoustic Imaging : Project Technical Report No.6*, Submitted to the Department of Electronics, India, June 1990.
2. B.Yegnanarayana, A.Ravichandran and R.Ramaseshan, "Transformation-invariant Object recognition from sensor array images," *Algorithms for Signal Reconstruction in Acoustic Imaging : Project Technical Report No.9*, Submitted to the Department of Electronics, India, December 1990.
3. A.Ravichandran and B.Yegnanarayana, "A two-stage neural network for translation, rotation and size invariant visual pattern recognition," *Proceedings of ICASSP '91*, Toronto, May 1991.
4. B.Yegnanarayana, R.Ramaseshan and A.Ravichandran, *Studies in Sensor Array Imaging*, Indian Institute of Technology, Madras, November 1991.

Other Publications

1. B.Yegnanarayana, R.Ramaseshan and A.Ravichandran, "Image reconstruction from multiple frames of sensor array data," *Indo- US Workshop on Spectrum Analysis in One and Two Dimensions*, November 1989.
2. B.Yegnanarayana, R.Ramaseshan and A.Ravichandran, "An algorithm for thinning noisy images," *Proceedings of ICASSP '90*, Albuquerque, April 1990.
3. B.Yegnanarayana, R.Ramaseshan and A.Ravichandran, "Improving resolution of sensor array images using multiple frames of data," *19th International Symposium on Acoustical Imaging*, Bochum, Germany, April 1991.