A STUDY OF SPATIAL DATA MODELS AND THEIR APPLICATION TO SELECTING INFORMATION FROM PICTORIAL DATABASES

BY

STEPHEN CHARLES

A doctoral thesis

submitted in partial fulfilment of the requirements

for the award of

Doctor of Philosophy of the Loughborough University of Technology

January 1991

This is to certify that the work presented in this thesis is original work undertaken by the author himself in the Department of Computer Studies, Loughborough University of Technology, along with the acknowledged references.

The work described has not been submitted in part or in full to this or any other institution for a higher degree award.

ACKNOWLEDGEMENTS

My thanks are due to Professor Ernest Edmonds, who made it possible for me to carry out this project, to Dr Steve Scrivener, who showed great skill and patience in helping to correct numerous problems and suggesting avenues of both study and research and to Mr Tony Clarke for his ideas on practical improvements to the thesis. I also greatly appreciate the help of all those at the Mac Consultancy Company in London and the Oracle Corporation at Belmont in the United States of America who both provided invaluable assistance with configuration of the laserdisk player and help with the construction of the relational database.

There is a rapidly expanding range of available spatial data in digital form, and a rapidly increasing need for its combined use.

(Donna Peuquet 1984)

ABSTRACT

People have always used visual techniques to locate information in the space surrounding them. However with the advent of powerful computer systems and user-friendly interfaces it has become possible to extend such techniques to stored pictorial information. Pictorial database systems have in the past primarily used mathematical or textual search techniques to locate specific pictures contained within such databases. However these techniques have largely relied upon complex combinations of numeric and textual queries in order to find the required pictures. Such techniques restrict users of pictorial databases to expressing what is in essence a visual query in a numeric or character based form. What is required is the ability to express such queries in a form that more closely matches the user's visual memory or perception of the picture required. It is suggested in this thesis that spatial techniques of search are important and that two of the most important attributes of a picture are the spatial positions and the spatial relationships of objects contained within such pictures. It is further suggested that a database management system which allows users to indicate the nature of their query by visually placing iconic representations of objects on an interface in spatially appropriate positions, is a feasible method by which pictures might be found from a pictorial database. This thesis undertakes a detailed study of spatial techniques using a combination of historical evidence, psychological conclusions and practical examples to demonstrate that the spatial metaphor is an important concept and that pictures can be readily found by visually specifying the spatial positions and relationships between objects contained within them.

CONTENTS

LIST OF FIGURES

Page no

CHAPTER 6

CHAPTER 7

CHAPTER 1

INTRODUCTION

Spatial data consists of points, lines, rectangles, regions, surfaces and volumes (Samet 1990). This data has been accumulating at an increasingly rapid rate over the past two decades representing a very major investment and an extremely valuable resource which is in demand for a wide variety of research and decision making applications, however attempts to integrate these data into existing systems have to date proven extremely difficult at best (Peuquet 1984). In the past databases have consisted purely of text, usually stored in a relational or entity-relationship format. However in recent years techniques for the storage, retrieval and manipulation of pictorial information within computers has become possible, due largely to the increase in resolution and processing power of the latest range of image display hardware, therefore the relatively low cost and increasingly high performance graphical capabilities of today's systems encourage users to develop and use databases which contain pictures. Such systems are commonly called pictorial databases.

A pictorial database can be formally defined as a collection of shareable pictorial data encoded in various formats. A pictorial database system provides an integrated collection of pictorial data for easy access by a large number of users (Chang 1981).

Initially these pictorial databases consisted of both textual information and images, however they were constructed by storing textual information separately from the pictorial data, linked only by textual registrations of the images. No attempt was made at detailed analysis of the information stored within the image. Examples include medical databases storing X-ray images which are linked to textual registrations of patient name and patient number (Assman, Venema and Hohne 1986).

Gradually systems using computerised digital analysis techniques were introduced in an attempt to automatically analyse the contents of the images stored within the pictorial databases. These systems have been further developed by the introduction of textual picture query languages such as GRAIN (Chang, Reuss and McCormick 1978) and IMAID (Chang and Fu 1980) which enable users to construct their own textual queries about the information stored within the images. IMAID for example is an integrated relational database system interfaced with an image analysis system. By using pattern recognition and image processing manipulation functions, simple pictorial descriptions can be extracted from images and stored in a relational database. User queries about pictures can then be manipulated through the relational database and pictures matching these queries displayed at query time thereby eliminating the need to process vast amounts of image data. Here the manipulation capabilities of conventional query languages are contained in those of the picture query languages (Chang and Fu 1981). If the user's request can be expressed in terms of the extracted picture descriptions, then there is no need to retrieve and process the original pictures. If on the other hand the stored descriptions are not sufficient, all pictures satisfying the given selection criteria can be retrieved from the picture store and processed until the required precision is obtained. Such systems provide a flexible technique by which objects and structures within a pictorial database can be analysed.

The relational model used within such systems has been amongst the most popular technique by which to analyse information within images. Within these systems relational databases store attribute values of objects and it is these values which are used to retrieve the pictures. This has been demonstrated to be a potentially powerful approach for handling complex queries for a broad range of applications, however the use of relational calculus to directly manipulate locational data has shown severe limitations (Peuquet 1988). The basic set operations of union, intersection and containment hold in a spatial sense, but this approach is derived purely from traditional mathematical concepts, there is no ability to handle inexact, context-dependent relationships, set-oriented or otherwise, or of defining higher-order relationships on the basis of simpler built-in operators. Thus there is a significant lack of knowledge about the spatial relationships within pictures stored in such systems.

It has also been demonstrated (Meier and Ilg 1986) that an extended relational database management system approach in which spatial relationships within a picture are directly encoded within a textual database is also severely limited. Such systems use a set of primitive textual spatial relationships when storing spatial data (Peuquet 1988). Here the designer views the spatial relationships between combinations of entities as another attribute within the database. Various lists of 'primitive' spatial relationships have been derived and stored as attributes but historically these seem to have been derived on the basis of intuition, rather than any fundamental theory or representational scheme. These 'primitives' include spatial relationships such as 'below', 'left-of', 'right-of', 'above' (Pizano, Klinger and Cardenas 1989). However the sheer number of relationships between different entities often makes the construction of such a system difficult.

It is therefore suggested that these text based picture query languages have several significant disadvantages which include:-

1) They are often difficult to learn and use, involving notations such as relational or predicate calculus (Chang 1981). An example of such a predicate calculus operation is:-

$$[x: \exists y \exists z ((y,z \text{ LAFAYETTE}) \in \text{CINAME} \wedge (y,w,x) \in \text{RONAME})]$$

which translates to:- display the names of the roads that appear in the same frame as city Lafayette.

2) They are usually domain specific, different application areas requiring both radically different textual picture query languages and image processing functions. An example of such a domain specific textual picture query language is GRAIN (Graphics-Oriented Relational Algebraic Interpreter). Here a system translator translates user queries specified in GRAIN into picture algebraic operations (Chang and Fu 1981). For example the user's query might be:-

paint highway; highway equal to '65';
within (city equal to 'LAFAYETTE')

This system then translates this query formulation into the following picture algebraic operations:-

```
t1=LTA(65)
t2=RESTRICT(pmt, "pict-object='LAFAYETTE'")
w3=WINDOW(t1)
t3=MASKO(t1,w1)
t4=JOIN(t3,LAFAYETTE,"x,y;x,y")
PAINT t4
ERASE t1,t2,t3,t4,w1
```

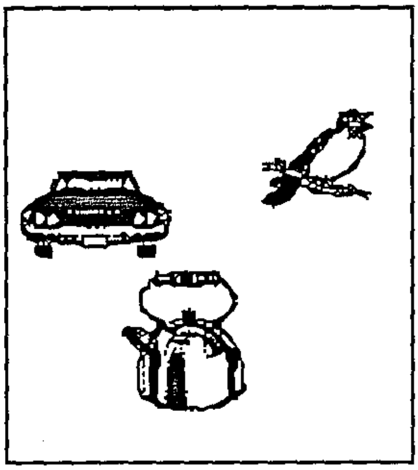
However within the user's query both highway and city are domain specific keywords which are inappropriate to other applications. In addition to this if information matching to the query is not stored within the relational database (highway 65 and LAFAYETTE, in the query above) then image processing functions will need to be applied to an image in order to find such information. The effectiveness of such image processing functions is therefore strongly dependent on the nature and form of the image being examined. This suggests that a more generalised image analysis technique is required which can be applied to greater variety of pictorial databases.

3) Statements in a textual picture query language need to be typed manually into the database management system, no spelling mistakes or ambiguous query terms are allowed. One such example is QPE (Query by Pictorial Example) developed by Chang and Fu in 1981 which allows the user to type relational terms into a table, however since the information is textual, exact spellings and relational combinations must be entered.

4) However one of the most serious problems with textual picture query languages is that information presented in pictorial form utilises the user's ability to rapidly register complex visual arrangements and relationships. Such structures are however extremely difficult to describe textually. For example using a relational database management system in which the spatial relationships are directly

textually encoded, figure 1.1 might be described textually as, car left-of bird, car left-of kettle, kettle-below-car, kettle-below-bird etc. Thus such data is stored in terms of topological relations rather than geometrical coordinates. These topological descriptions are generalisations, since in fact for figure 1.1, the car is not only left of the bird but is also inclined at an angle to it. Users may thus be restricted in their ability to find such a picture based upon such generalised textual descriptions (many pictures may contain a car to the left of a kettle).

FIGURE 1.1



A PICTURE WHICH THE USER WISHES TO LOCATE FROM A PICTORIAL DATABASE

In this thesis it is proposed that for pictorial databases, a possible solution to this problem is an interactive spatial database system, where users can spatially indicate the nature of their query and then have the system identify pictures in its pictorial database which match these requested spatial configurations. This concept relies upon both the user's and system's ability to identify the spatial arrangements of the objects contained within the pictorial database. (Spatial here is defined as: 'of pertaining to, or relating to space' (Clark 1985) wherein space is not based on an interpretation of the properties of things but from the very start based on actions and by itself an action that is performed on the things (Piaget 1975).

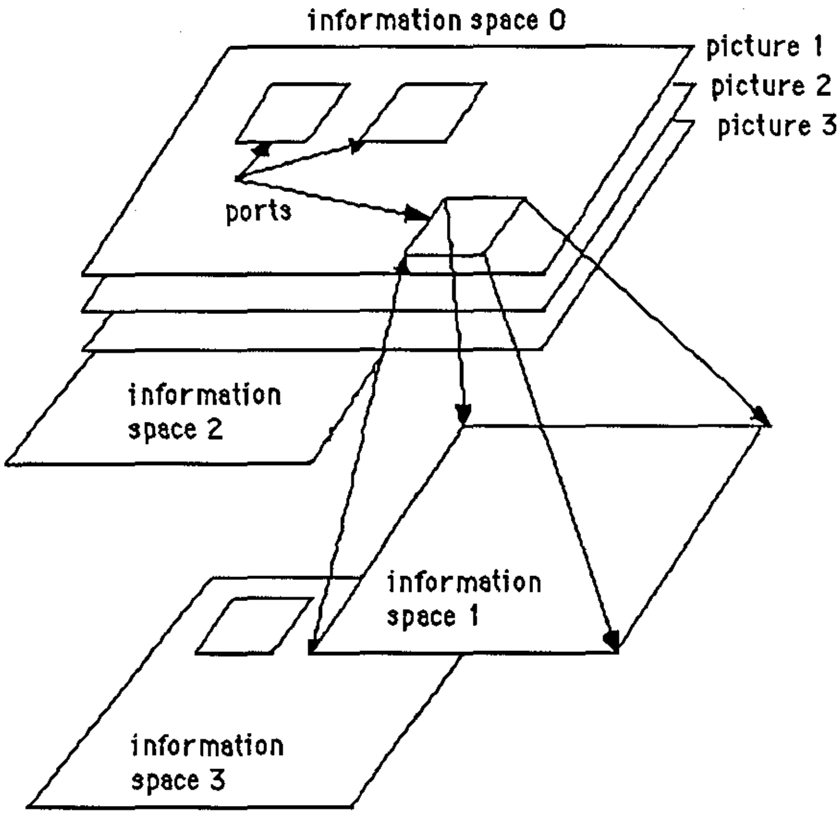
"People use spatial methods of locating information naturally in their daily lives, usually without realising it. The ability to recall where on one's bookshelf a particular book resides, is due at least in part to stored spatial memory of the location of the book irrespective of the subject matter within it. To be complete, location as a clue to finding data is augmented by iconic appearance. Thus location gets a subject close to the book required, but its tall skinny blue binding (or other attributes such as its size or shape) helps to locate the exact book being searched for" (Fields and Negroponte 1977). Therefore we understand space by moving ourselves, things and data through the space (Kohler 1987).

It has been shown (Calloway 1989) that when spatial metaphors and artificial realities are used to organise human computer interactions for decision making, users are subject to the same learning constraints and opportunities as they are with normal spatial realities. Since normal spatial realities have been extensively examined and subjects found to have excellent spatial memories for such interactions (Evans and Pezdek 1980), this suggests that these spatial metaphors may be fruitfully applied to artificial realities.

Existing pictorial databases which use such artificial reality methods of locating information include Herot's spatial database of ships, Lippman's movie map and Bolt's dataland system. Herot's database of ships (Herot 1980) allows the user to navigate through a pictorial database in order to find information about a variety of American and Russian ships. Lippman's movie map (Lippman 1980) uses a spatially interactive map to enable the user to undertake a simulated drive around a small town in America. Bolt's media room at the MIT's Architecture Machine Group laboratory (Bolt 1984), is a setting where the room itself is the terminal. This room (which contains a system called Dataland) allows users to interactively move through virtual space in order to select windows from a screen displayed in front of them. All three of these systems allow users to navigate through a virtual three-dimensional information space in order to find the information they require from a pictorial database, usually by direct manipulation using a mouse. The concept of three-dimensional virtual space is demonstrated in figure 1.2.

Thus pictures are found by remembering where within the virtual three-dimensional space they are located rather than the identification of such pictures based upon their names or other attributes. Systems allowing users to navigate through virtual three-dimensional space are called Spatial Database Management Systems (SDMS). Such systems allow users with little or no knowledge of the applications area to readily find pictures contained within them. Here each information space may contain several pictures however selection of a picture from within one of these information spaces enables pictures within other information spaces to be displayed. These other information spaces are said to be connected by ports and often contain more detail about the picture initially selected.

FIGURE 1.2



VIRTUAL THREE DIMENSIONAL SPACE

1.1 THE THESIS PROPOSAL

The proposal in this thesis differs from this work in that rather than using spatial methods to navigate through a conceptual three-dimensional space in order to locate information from a pictorial database, the aim is to examine the value of making explicit use of the two-dimensional spatial layout of objects within pictures. The incorporation of such a technique into a database management system may then allow a picture to be searched for by recall of the spatial configurations of objects within it.

Identification of pictures based upon the spatial configuration of objects within them relies upon the mental imaging and perception skills of the user. Imagery here is the ability to form a mental picture or image from which more information can be obtained, while perception is the ability to detect structures and events in one's surroundings (Palmer and Frank 1988). A perceptual and reasoning ability is of increasing importance in the capability to interpret and analyse imagery, to detect objects and to notice changes in pictures (Tranowski 1988). It has been proposed that there are at least three kinds of memory-relevant information a subject could extract from a picture which might be influenced by the nature and organisation of that picture (Mandler and Johnson 1976). Firstly the objects in the picture constituting the scene, secondly the physical appearance of the objects in the picture and thirdly the actual location of the objects as well as the spatial relationships between them.

Since subjects appear to remember spatial relationships between objects (Hasher and Zacks 1979), then it is suggested that it should be possible to locate pictures from a pictorial database by recalling the spatial configurations of objects within these pictures. For example in the case of figure 1.3 it is suggested that such a picture can be found from within a pictorial database by the specification of the spatial arrangements of the objects within it. Having specified the spatial arrangements of the objects on an interactive spatial interface in the approximate positions that the user remembers them, the database management system then attempts to match the information within the interactive spatial interface to information stored within the pictorial database. Thus for example in figure 1.3 a user may consider that two of the most memorable objects were the glass and the scissors and therefore placing representations of these objects on the interactive spatial interface in appropriate positions should enable matching pictures (such as figure 1.3) to be found from a pictorial database. Possible applications for this work include:-

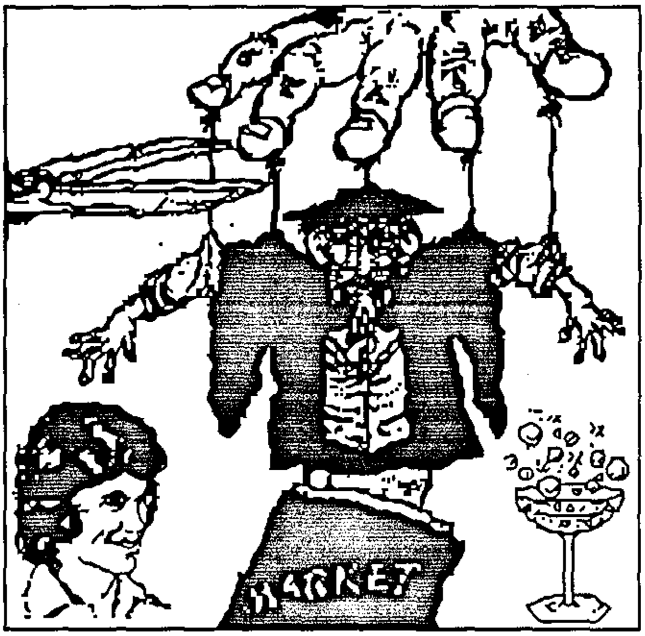
1) Star Maps

Here user requested spatial arrangements of stars (such as the plough) could be entered on an interactive spatial interface and the database management system could then locate a picture with a matching spatial configuration from its pictorial database of star maps.

2) Geographical Applications

Geographers are often interested in particular spatial layouts of geographical information within terrain maps in order to indicate suitable locations for their work. For example all possible locations of hills to the north of lakes may be of interest to a geographer looking for the best location for tree planting (Chang, Yan, Dimitroff and Arndt 1988).

FIGURE 1.3



A TYPICAL PICTURE WHICH CAN BE LOCATED BASED UPON THE SPATIAL ARRANGEMENTS OF THE OBJECTS WITHIN IT

3) Military Applications

Military applications include interactive spatial systems which contain the location of military personnel and weaponry in order to plan the best spatial layout for engagement with an enemy. Here the nature of the photo-interpretation problem and the requirements of knowledge acquisition require that the knowledge is visually oriented and that the relationships between objects can be described. (Tranowski 1988).

4) Art Applications

Works of art by individual artists often have particular spatial arrangements within them which characterise their work (Vaughan 1988). Spatial analysis of these works of art may therefore help in their classification.

Thus to conclude the main proposal within this thesis is that:
An important attribute of pictures is the two-dimensional spatial arrangements of the objects contained within them and that this attribute can be used to locate specific pictures from a pictorial database. A number of assumptions underlie this proposal, these are:-

1) The two-dimensional spatial layout of objects within a picture is often visually memorable and therefore a system which enables the user to recall information based upon its spatial layout is a possible technique by which information may be selected from a pictorial database.

2) There are a number of pictorial database applications in which the two-dimensional spatial layout of the objects contained within these applications can make a significant contribution in the analysis of the stored information.

3) That it is feasible to use a computer system to both store and access pictorial information and the accompanying database management system.

This thesis uses current spatial systems, psychological evidence and practical demonstrations in order to show that using a spatial method of locating pictures based upon the two-dimensional spatial configurations of objects contained within them is feasible. The history of both computerised and non-computerised pictorial classification will also be discussed, demonstrating previous techniques for locating and analysing pictures. The structure of the thesis is as follows:-

Chapter 2) Non-Computerised Techniques
This chapter outlines the methods used in the classification and sorting of pictorial information before the advent of computers. An appreciation of these techniques

is important because there is the possibility that such principles may be applicable to the analysis of pictorial information stored within today's computers. Today's image manipulation techniques and computerised classifications of pictorial information can be seen as a logical extension of these earlier photographic methods.

Chapter 3) History of Databases

This chapter defines commonly used database terms and shows how databases have developed from those storing simple textual data to those containing highly sophisticated pictorial information systems. These systems are important because they were the precursors of the more specialised spatial database systems and as such need to be discussed in this context.

Chapter 4) Spatial Databases

This chapter describes databases in which data is selected based upon its spatial arrangement usually within three-dimensional virtual space. It demonstrates the various practical applications of such systems and their advantages and disadvantages when compared with non-spatial pictorial database systems.

Chapter 5) The Psychology of Visual and Spatial Techniques

This chapter discusses the psychological importance of both picture identification and spatial analysis of visual information. It attempts to show the various theories concerning the nature of imagery, memory, perception, recognition and shape analysis and the application of these theories to pictorial databases. Such a study is important because these theories attempt to explain how we both view and interact with the world around us and therefore the more effectively systems can be designed which emulate these real world mechanisms then the greater the potential productivity in using such systems.

Chapter 6) Proposals for the Analysis of a Pictorial Database Based in Part upon a Spatially Depicted Search

This chapter uses both the evidence from psychological studies and pictorial database systems to propose a database system in which users can locate pictures by depicting their spatial query. This is achieved by selecting one or more pictorial representations which are then placed on the user interface from which pictures containing matching spatial information can be found. It also discusses the conditions under which these techniques can be applied and gives practical examples of their use.

Chapter 7) A Practical Pictorial Database Classified by Different Visual Attributes
This chapter uses the conclusions from the previous chapter to construct a database management system in which spatial position and other visual attributes of objects are used in locating pictures. It also suggests extensions and problems with the system.

Chapter 8) Conclusions
This chapter attempts to draw conclusions about the spatial metaphor as a technique in the locating of information from pictorial databases.

Chapter 9) References

It will therefore be argued in this thesis that the two-dimensional spatial arrangement of pictorial information is a valid and useful way of constructing descriptions in the analysis of a pictorial database. For pictorial databases where only a small number of similar objects are present it will be argued that this is one of the most effective techniques by which such data can be analysed. However even in the case of more generalised pictorial databases containing a much larger number of more visually varied objects, then data associated with spatial attributes of the objects when combined with other object attributes such as name, colour, size and orientation can still provide a considerable contribution in both the analysis and retrieval of pictures present within the pictorial database.

CHAPTER 2

CLASSIFICATION AND STORAGE OF NON-COMPUTERISED PICTORIAL INFORMATION

Section no

2.1 INTRODUCTION

2.2 INFORMATION RETRIEVAL SYSTEMS

2.3 EXAMPLES OF NON-COMPUTERISED PICTORIAL CLASSIFICATION SYSTEMS

2.3.1 Simons and Tansey's Classification Scheme

2.3.2 Shell Photographic Library Classification

2.3.3 Alphabetic Subject Arrangement

2.3.4 Place vs Subject

2.3.5 Personalities

2.3.6 Geographical Names

2.4 CLASSIFICATION OF MAPS

2.4.1 Non-Geographical Classification

2.4.2 Geographical Classification

2.4.3 Numerical Geographical Classification Schemes

2.5 CONCLUSIONS

2.1 INTRODUCTION

Pictures are a source of visual information. The information contained within a picture may be taken from life, geographical features, scenery, street scenes, natural phenomena, animals, people or activities, even a work of art. Whatever the nature of the pictures there is a requirement for their classification and storage.

Pictures can take many forms, the term 'picture' may be used to mean anything from a cave painting or an original work of art, to a film frame or photograph. Within this chapter only photographs and maps will be referred to, however the principles used can be applied equally to almost any form of pictorial data.

Libraries have for many years been storing and analysing non-computerised pictorial information, and therefore a study of the procedures they use is helpful in that computerised methods for both storing and analysing pictorial data can then be seen in context. Libraries usually store pictorial information in the form of photographs, and here two of the most difficult problems has been the initial classification of the photograph and the ease by which a person can locate it from the library collection. Often the information required has been retrieved from collections of photographs simply by searching through them individually in a sequential fashion. However since many library picture databases are very large, an important issue is that of how the designer of these systems can help the user in the formulation of their queries (Bennett 1971).

2.2 INFORMATION RETRIEVAL SYSTEMS

The term Information Retrieval refers to the process by which information is organised for the purpose of searching, and the processes by which the resultant information is searched (Booth and South 1982). In a sense many documents, parts of documents, or pictures within a library, are themselves information retrieval (IR) systems (for example the alphabetical index of a book can be seen as an information retrieval system), but the term is usually reserved for those documents or systems specifically arranged for searching such as dictionaries or descriptions of pictures. Information retrieval systems within libraries are usually designed to perform one or more of the following functions for users (Booth and South 1982):-

1) Reference Retrieval. Here the library system provides references to documents or pictures in which information may be found.

2) Document or Picture Retrieval. Here the library system produces the documents or pictures in which information may be found.

3) Data Retrieval. Here the library system produces the information required without an intermediary, therefore the system directly displays the data required rather than taking the user to the document heading, summary or picture. Thus in the case of a picture file an object within a picture is displayed rather than the entire picture.

Documents or library files need in general to be created for one or more of the following reasons. Firstly the prevention of loss through damage or deterioration, (the preservation function), secondly the prevention of loss through unauthorised use (the security function) and thirdly the prevention of loss through dispersal where a document is lost because the retrieval system is incapable of locating it within the system (the search function).

Of these three reasons the third is the most important within this thesis. A file (which can be defined as a collection of documentary items arranged for the purposes of searching (Booth and South 1982)) contained within the library classification system must be capable of supporting either of the following types of search inquiry specified by the user. Firstly search for an item (or group of items) which have been previously selected by the user and are known to be in the system, and secondly a search for items which are not known to be in the system. In the second case where items are not known to be in the system then there are several possible reasons for searching. To confirm that the items exist, to confirm the non-existence of items, or to ensure that all items in a certain group are found, thus finding out 'how many' or 'how much'.

In all cases where more than one item is being sought, the ultimate purpose may be to retrieve a group of items from which a further selection is to be made. In the first instance such a file (whether it be of text or pictures) may be totally unorganised. The items within the file may be in an arbitrary order which may change every time the file is used. Given such a disorganised file it will be necessary to search it item by item until the items sought are found. In the case of the complete retrieval type of inquiry in which all items concerning a specific subject are required then this will always entail searching the whole file. However such a file is not necessarily ineffective and indeed it may be more effective than an organised file. If the items in the file are relatively few in number, or if they are easily recognisable and/or distinguishable from one another, then there may be little point in using time in organising the file in which the information is stored.

There is a point however at which some kind of pre-organisation becomes necessary. This is the point at which the repetitive effort of searching the whole file becomes unacceptable, or when the time taken to respond to an inquiry becomes too long. Typically this point will be reached when the items in the file become too numerous. The extent to which it is desirable to pre-select items within the file, depends upon frequency of use, desired response times, and complexity of the information stored within the file. It is therefore important to make proper decisions regarding the criteria to be used in the creation of classes in such a file. This is particularly true for files of pictures. Within libraries the following general principles are applied for the storage of pictorial information.

1) Things which are in some way connected or related are often stored together, for example pictures of books can be stored together with other pictures of books, an alternative technique is to store pictures of objects which are commonly used with each other (such as knife and fork) together.

2) The classes should be definable by objective criteria. Nice looking cars and nasty looking cars are real classes, but these can not be incorporated in main filing systems because essentially they are classes which involve a subjective description of the objects (what one person perceives as a nice looking car, another person may perceive as a nasty looking car) and therefore filing information based upon these subjective classes would lead to an unorganised picture file system.

3) Membership of individual classes should be small enough to limit the choices at each search to a convenient number and to allow for easy visual scanning of the pictures present.

The characteristics by which classes are defined, and about which questions will be asked are called the sought characteristics. When these characteristics are turned into labels they become 'sought terms' and because the sought terms to a greater or lesser extent determine order (the way in which items are sorted) they are also called 'sort terms' (Booth and South 1982). Any characteristic of a document or picture can be a sort term and it follows that a document or picture can belong to more than one class. In most systems it is necessary to limit the characteristics used in classifying items, and it is therefore important to determine which are the essential characteristics, upon which questions are most likely to be asked. The criteria that have been found most useful for classifying pictures within libraries relate to the following:-

1) Content: animals, cars, rivers, etc.

and secondly

2) Criteria affecting content:

Bias :Who is the photograph designed for?

Media: What form does the information take?

Type of use: How is the resulting information to be used?

Date the photograph was taken.

Many ideas and concepts such as beliefs and theories are not represented in a pictorial form, because they can not be easily symbolised (Wright 1981).

When categorising pictures it must be remembered that some forms of classification may be not possible for all cases, for example categorising a photograph by author is not possible where the photographer is unknown. Pictures are also difficult to describe, the title of a picture does not uniquely identify that picture (Kosslyn and Chabris 1990). Since pictorial classification is not without a degree of difficulty, several different schemes have been developed for the storage and classification of non-computerised pictorial information.

2.3 EXAMPLES OF NON-COMPUTERISED PICTURE CLASSIFICATION SCHEMES

2.3.1 Simons and Tansey's Classification Scheme

One such scheme is due to Simons and Tansey (1970) at the University of Santa Cruz, California. Their scheme is a textual method for organising and automatically indexing interdisciplinary collections of slides and pictures. This system is a hierarchical classification becoming more specific as the search space is reduced, however it is very heavily biased towards classification of art photographs and pictures and has numerous subdivisions of fields within the arts domain. The classification technique is based upon several assumptions:-

1) The collection and the classification should be general, encompassing the subject matter of all academic disciplines.

2) The arrangement of the collection should reflect a broad historical and cultural approach to teaching (which is its main application area).

3) The filling arrangement of the collection should encourage and facilitate browsing in the files containing the classification scheme.

4) The collection should be fully catalogued or indexed, preferably by automatic means.

Within Simons and Tansey's classification system three main subject divisions are used:- Art (which includes all man-made artifacts), Science (which includes natural phenomena and scientific principles) and History (which includes the rest). Fields used within this classification include Chronological Period, Country, Subject, Title and Primary and Secondary Key words. For a picture of trees being cut in the Amazonian forest, the classification using Simons and Tansey's technique might be:-

G848E.D

whereby
G represents the 20th Century (within History)
848 the Amazon Basin
E represents Forestry
and D represents Tree Felling.

However although Simons and Tansey's classification system has been extensively applied to non-computerised pictorial information, in practise one of the problems is that there is no way in which it can specify people in general rather than as people in occupational roles. This is due to the subdivisions of subjects within the classification, objects outside of the three main headings of history, art and science being difficult to classify.

2.3.2 Shell Photographic Library Classification

A different picture classification system has been used at the Shell Photographic Library. Here a coordinate indexing system was developed in which a number of identifiers were used to describe the characteristics of the picture in question (McNeil 1966). These identifiers in the form of keywords were used to index the 60,000 photographs present. The photographs were sorted according to the main areas of activity within the company these areas being

G-General-Matters, H-Historical, K-Manufacturing-Chemicals, L-Marketing-Chemicals, M-Manufacturing-Oil, R-Research, S-Marketing, T-Transport and Storage, X -Exploration and Production.

On average 15 keywords were required within this system to describe a picture, the range being between 5 and 30. These keywords could be used to describe any aspect of the picture, however descriptions based upon spatial information was not included. Unfortunately this keyword method of description was found to be too detailed for most users at Shell. For example users in the publicity department usually made general enquiries about pictures concerning a topic, rather than a specific event or object. In summary the conclusions at Shell about coordinate indexing was that:-

1) The facility to browse was sacrificed.

2) There was a tendency to build up the pictures into an index when being classified. This index contained a large number of keywords used to describe each picture. The storing of these numerous keywords in the form of an index made retrieval slower.

3) Costs were very high when detailed indexing was carried out.

4) Coordinate indexing was most useful for small specialised collections where a great deal of detail was required.

The major advantage of the Shell system was that because of the large number of keywords used to classify pictures, users were able to locate the pictures they required based upon many different descriptions of the information contained within them. Therefore providing that the picture in question could be contained within one of the main areas of activity, then there was a large degree of certainty that it could be found. Photographs outside Shell's main activities were however difficult to classify. This led to the category of General Matters (containing activities outside of the Shell company) becoming artificially enlarged with pictures that could not be easily classified elsewhere.

In addition to these two commercial forms of picture indexing many other forms has been used, the most important of which are discussed below (Wright 1981).

2.3.3 Alphabetical Subject Arrangement

Since most library staff are not trained in methods of pictorial classification, pictures are often classified alphabetically based upon the most significant item present within the picture. However this leads to problems in that large groups of objects under the same alphabetical heading are constructed, such as birds or flowers. Thus if an alphabetical arrangement is adopted, then sub-divisions are usually required under the main heading. Therefore under the general heading of birds, sub-divisions might include birds of prey, sea birds, wading birds, flightless birds, etc. The major advantage of this system is its ease of use. Sorting pictures alphabetically is probably the easiest method of classification where the picture contains a single well defined object. However in the case of pictures which contain several objects, classification becomes more difficult. This is because of the problems as to which of the objects present within the picture should be included in the classification.

2.3.4 Place vs Subject

This method of picture classification relies upon the ability of the librarian to classify pictures under the headings of either PLACE or SUBJECT. Using this method non-UK subject pictures are classified under PLACE giving rise to such headings as, France, Spain, Germany etc. Example items under the SUBJECT heading include Agriculture, Army, Education, Law and Police.

The problem with this classification system is that it is sometimes difficult to decide whether to classify a picture as a PLACE or a SUBJECT. Entertainment, Science, Religion and Sport can all be classified under either PLACE or SUBJECT since for these activities either may be important within the picture. For example within Entertainment, bull fighting may be classified under SUBJECT, or since it is most famous for occurring in Spain, it may be classified under Spain, which in turn would be classified under PLACE. Therefore the classification of headings under these circumstances is purely subjective and ultimately dependent upon the person classifying the picture. However for well defined headings such as place names, or well defined subjects such as physical objects then this method of classification is one of the most effective.

2.3.5 Personalities

Classifications of pictures by personality are usually arranged in alphabetical order (Harrison 1981). However sporting personalities are a problem because action shots containing well known sportsmen tend to be asked for separately. For example a picture of Nigel Mansell in his formula one Ferrari could be classified under personalities (because his car is immediately recognisable) or could alternatively be classified under racing cars. In many cases such a distinction is difficult. Therefore this method of classification is only ideally suited for photographs of personalities which do not contain any reference to the activity for which they became famous for. Therefore under this form of classification a picture of Nigel Mansell would not contain a picture of his racing car, because this could lead to an alternative method of classification producing an unsystematic system.

2.3.6 Geographical Names

Geographical Names are the names assigned to geographical locations. Such a method of classification is only really suited to the sorting of geographical data where the user may require the names of all towns, cities or countries within the pictorial database (Harrison 1981). However the names of the locations may change, either by changes in the name of the country or town, or in the boundary definitions.

2.4 CLASSIFICATION OF MAPS

Geographic pictorial data have traditionally been presented for analysis by means of two-dimensional analog methods known as maps (Board 1967). Map classification can be broadly divided into non-geographical and geographical techniques (Larsgaard 1978).

2.4.1 Non Geographical Classification

Non geographical classification is usually carried out by one of the following identifiers:-

1) Subject
2) Date
3) Accession Number
4) Provenance
5) Size

Here the form of classification may not be logical and in many cases is extremely poor (maps have even been classified by their physical size!). Under such circumstances unless the user of the system already has a good knowledge of the maps available then finding the correct map will be difficult. In the case of classification by size, the library may have a 6 metre square map of the Isle of Man and only a 1 metre square map of the Soviet Union, therefore classifying by size would place the map of the Isle of Man near the end of the classification sequence and the map of the Soviet Union near the front. Thus the user searching for the map of the Isle of Man would need to search through nearly all of the maps in order to find the correct one. Under these circumstances unless the user already knows the approximate size of the map, then a serial search through all the maps would need to be carried out each time an individual map is required.

2.4.2 Geographical Classification

There are three basic types of geographical classification within maps:-

1) Alphabetic
2) Alphanumeric
3) Numeric

Alphabetic classification simply classifies each of the geographical areas of interest in terms of a textual abbreviation. For example

Afghanistan	AF
Albania	AL
Algeria	AG
Andorra	AN

Alphanumeric classification is the most popular technique by which maps are classified. However within alphanumeric classification there is a multitude of different classifications, only one of which is discussed here, the US Library of Congress Classification System, (Schedule G). This system is considered the best classification system for maps and is capable of a large degree of expansion and compression. Within the US Library of Congress Classification System world globes of various sizes are classed G 3160 to G 3182 and maps are given the numbers G 3190 to G 9980. Each major cultural or political unit of the world is assigned a block of numbers, the listing consisting of the letter G followed by a four-digit number. Each block of numbers has the endings either 0 through to 4 or 5 through to 9, therefore producing a range of five numbers for the block. 0 and 5 are general maps, 1 and 6 are subject maps, 2 and 7 are regional maps, 3 and 8 are county maps and 4 and 9 are city maps.

Classification headings include:- Universe, World, America (with subdivisions), Europe, etc.
Particular types of maps are further classified in terms of their content. For example:-
Special categories, Mathematical geography, Physical sciences, Human and cultural geography, Political geography, Economic geography. These are themselves further divided. Often where possible the date of publication of the map and authority responsible for the map are also included in the classification. For example for a geology map of Washington state, the complete classification would be:-

G4281	Washington State
C5	Geology
1961	Date of publication
.U5	USGS (the authority producing the map)

Other alphanumeric map classification methods such as those of Boggs and Lewis, Dewey Decimal, and American Geographical Society all use similar techniques, the only significant difference being in the actual number and letters used in the classification of the areas of interest. These alternative methods of alphanumeric classification bear a close resemblance to the classification of textual information within libraries, in that they both use arbitrary alphanumeric characters in their classification schemes.

2.4.3 <u>Numeric Geographical Classification Systems</u>
These classification schemes have recently received an upsurge in popularity. One such example is the Universal Decimal Classification (UDC). UDC attempts to divide the whole field of knowledge into ten main classes denoted

by decimal fractions .0 to .9, .9 representing geography. Again as in the case of alphanumeric classification, further subdivisions are possible depending on the nature of the map, its contents and scale. By convention numbers are separated by parenthesis.

For example (63)(58.19)(84.3)

(63) representing Ethiopia
(58.19) representing Plant geography
(84.3) representing maps
Thus such a classification represents a vegetation map of Ethiopia.

2.5 CONCLUSIONS

An individual picture may in general be classified by its title, authorship, edition, publisher, date of publication or issue, size, physical form, illustrative matter, relationship to another document, bibliography, intended readership, or membership of a named or number series. Some of these classifications are unique identifiers, others when used in combination with each other become more specific such as title and author, title and series etc. The requirement to organise pictures depends on such factors as their frequency of use, the response time required and the complexity of information stored.

Booth and South have suggested that the following types of search are important within information retrieval systems such as pictorial database management systems. Firstly search for an item (or group of items) which have been previously selected by the user and are known to be in the system, and secondly a search for items which are not known to be in the system. In the second case where items are not known to be in the system then there are several possible reasons for searching. To confirm that the items exist, to confirm the nonexistence of items or to ensure that all items meeting a certain requirement are found (such as how many) are found.

Non-computerised pictorial classification has both advantages and disadvantages in the analysis of pictures. Its advantages are that because of their non-computerisation, these systems are easy to develop once the principles involved are understood. There are a large variety of non-computerised systems available, usually specialising in a particular application area. Users need not have a detailed knowledge of such systems in order to classify and locate pictures. Another advantage of non-computerised methods is that physical rearrangement and removal of pictures from the database can be easily carried out.

However there are several disadvantages. One of the past problems with the indexing of visual materials is that the characteristic of 'aboutness' does not exist for pictures in the way that it does for words. Textual interpretations of pictures differ from person to person. For example for a picture of a seaside scene, whereas one person may interpret the theme of a picture as being that of a beach, another person may interpret the theme of the scene as being holidaymakers. This leads to significantly different classifications which seriously hinders retrieval of the picture by other users. Thus there is a lack of standardisation by which subject indexing of images may be guided. Because no formalised standards for picture classification have evolved, designers tended to invent their own systems. These classifications have lead to a somewhat arbitrary ordering of pictures, largely dependent on the person classifying the picture. Such a classification may also be simplistic.

An ideal system would hold a complete record of all relevant information, and should allow retrieval based upon either individual or combinations of attributes within the system. Another disadvantage is that none of the non-computerised systems discussed are ideally suited to generalised databases in which a number of varied objects need to be classified, each method tends to specialise within a restricted application area.

Yet another disadvantage is that because of the non-computerised nature of the pictorial databases (information being in the form of photographs, film, slides etc), there are considerable problems with aging, decomposition and loss of information, searching such information (with or without an index) is also time consuming manual search being required.

In summary there are many different kinds of classification used in the sorting of non-computerised pictorial information. Systems discussed within this chapter however are usually applied to only narrow application areas (such as map classification or Simons and Tansey's system which subdivides pictures into history, art and science). This means that their applicability to a more generalised pictorial database, containing a wider variety of objects is somewhat limited. Also none of the systems discussed makes any attempt at the analysis of spatial data contained within the pictures. However such systems are quite adequate for the storage of pictures within their own restricted application areas especially where a rapid accessing time is not required.

It is suggested that users looking for a classification system for non-computerised pictorial information should first search the available methods in an attempt to see if a system exists which encompasses the application area of their interest, since there are a multitude of systems which are specifically designed for such restricted application areas. However in the case of pictures containing a greater variety of objects then the choice of system depends to a large extent on the user's own personnel preference, since each system has its own merits and deficiencies.

Clearly there is a need for specific information to be located from a pictorial database. However non-computerised classification and retrieval techniques appear to be restricted in their ability to allow users to easily locate such information. Computerisation techniques have to a large extent superseded non-computerised classification and this is due primarily to their ability to rapidly sort, search and retrieve the pictorial information, although the actual techniques used to locate the information still suffer from many of the problems associated with non-computerised techniques. This is further discussed in the next chapter.

CHAPTER 3

DATABASE HISTORY

Section no

3.1 INTRODUCTION

In order to understand the origins, principles and deficiencies involved in the analysis of pictorial databases, conventional textual methods and their extensions into the pictorial domain, need to be understood. Therefore within this chapter the theories and methods used in these textual database structures are explained and constructively criticised. The progression from these systems into systems analysing purely pictorial data is then examined and how such pictorial systems attempt to analyse spatial information. All of the following methods within this chapter use computerised techniques to access, store and display information, usually in the form of a database management system specially constructed for that purpose.

3.2 DATABASE DEFINITIONS

A textual database can be defined in several ways, these include:-
As a collection of non-redundant data shareable between different application systems (Howe 1983) or alternatively as a file of data so structured that appropriate applications draw from the file and update it, but do not themselves constrain the file design or its content (Chandor 1977). Both of these definitions underline the fact that the main function of a database is to inform the user or users about information stored within a computer system. In general the purpose of a database is to store all of the data in one location such that redundant data is eliminated. This is because when multiple copies of data are stored at different locations and each updated individually, then at any one time two different versions of the same information may be present, making it impossible to determine which is the correct set of data. Therefore within a well designed and structured database these discrepancies are eliminated.

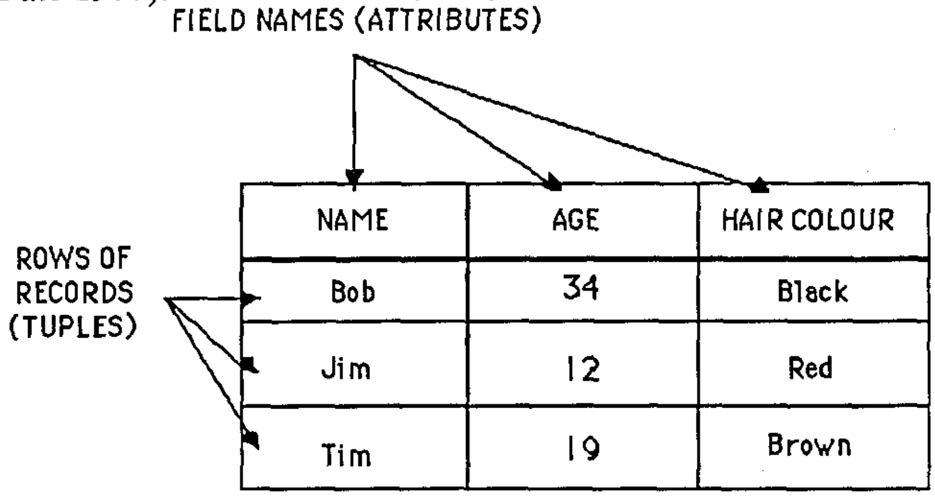
Depending on the size of the computer, the information held within a textual database may be accessed by one user or by many hundreds of users throughout a distributed network. Storage capacities vary from many thousands of megabytes on mainframe computers down to as small as a few kilobytes on a microcomputer. All textual databases, whether on a large mainframe or a small microcomputer have a structure which enables the user to access and manipulate textual information stored within them. Within both mainframes and microcomputers one of the most common approaches to the design and use of such databases has been the relational approach.

3.2.1 The Relational Approach

A relational database is a collection of relations that contains all the information that is to be stored in the database (Jackson 1988). Within a relational database the smallest unit of data that has meaning to its users is called a field. The fields about a particular item are grouped together to form records, and a program usually reads or writes whole records at a time. Therefore each record usually contains all the information about a particular item such as its name, address, sizes or colours etc. The name of these fields of information are said to be the attributes of that particular item. Within relational database nomenclature, items are more often called entities.

An entity has various attributes about which it is wished to record information. Usually a record is maintained about each entity (item). Similar records are themselves collectively grouped into files. The column titles within a relational table are the attributes or fields. Each row within the database is called a record or tuple and contains all the information about a particular item. All the records in the database are combined to form the file, alternatively called the table or relation (Date 1977).

FIGURE 3.1



A SIMPLE TEXTUAL DATABASE FILE CONTAINING THREE RECORDS AND HAVING FIELD HEADINGS OF NAME, AGE AND HAIR COLOUR

3.3 GRAPHICAL DATABASE INTERFACES

Initially most databases used the relational database approach when storing textual information. In order to make such databases easier for users to work with, more complex user interfaces were developed. An interface under these circumstances can be regarded as a notional boundary or surface that joins together a computer and the user (Barker, Najah and Manji 1987). Many of these user interfaces employed graphical methods of representing information which employed the user's capacity for spatial perception and analysis.

Such graphical interfaces often helps the user's comprehension and interpretation of the information they have requested. They can also be used to enhance the initial query formulation by using a graphic language which enables the user to work with some form of diagram. These graphical representations therefore take advantage of available interfaces and pointing devices (such as mice) to provide a friendlier user-interface to a database system. This benefits both computer novices and experienced users. If a system is infrequently used users may have difficulty in obtaining sufficient operating skill to successfully achieve their goals. Graphical database interfaces are easier to learn and usually more memorable than data displayed textually, thereby aiding recall during their use (Veith 1988).

There is also evidence to suggest that graphical communication methods based upon the use of such facilities as windows (Norman, Weldon and Shneiderman 1986), icons (Gittins 1986) and comics (Kindborg & Kollerbaur 1987) can provide powerful and efficient mechanisms for the facilitation of human-computer interaction enabling the users of such systems to more readily access and extract the information that they require.

3.3.1 Query by Example

Among the first to use graphical methods to analyse relational databases was Zloof (1975), who designed a system in which the database was queried by placing an example of an answer within the question. This query by example language was an attempt to appeal to the non-professional user and operated on a relational model of data as introduced by Codd (1970). In this system users formulates their queries by 'filling in' the appropriate table rows with an example of a possible answer. Thus for a large class of 'simple' queries users needed to only distinguish between the example element or variable (which is underlined), and the constant element (which is not underlined). Such information could also be printed by placing a capital letter P before any data that is to be output.

For example for a query to print the red items within a textual database, then the user fills in the table provided in the following manner.

FIGURE 3.2

TYPE	ITEM	COLOR	SIZE
	P.PEN	RED	

QUERY BY EXAMPLE

Since the query is concerned with red items, RED is a constant element and is therefore not underlined. On the other hand the underlined element PEN is referred to as an example element, and is entered as an example of a possible answer. Zloof's work was amongst the first of a number of systems whose design reflected a growing awareness of the importance of graphical methods in the efficient use of databases.

Some of the unique features of this system were that:-
Firstly the user perceives the interaction as manual table manipulation. This means that the user registers an impression of direct manipulation of the database through query construction by table filling.

Secondly the user can easily pre-identify the relations to be used, resulting in a early reduction in the scope of the database search. The required format of the answers can therefore be chosen by the user so that each answer only displays the information that are of interest to him/her, not all of the information concerning every item within the database.

Thirdly there is a preestablished frame of reference. The user views a table after having entered an example of the answer that they require thereby ensuring that they understand the form that the answer will take.

Fourthly as opposed to linear-type languages where the user is constrained to one degree of freedom, here the user has multiple degrees of freedom in that the sequence of filling in the tables and the rows within the tables is unconstrained. The constant elements, variables and print specification can also entered in any order. Under such circumstances the user's thinking process is less constrained during the formulation of a query.

Zloof's system was amongst the first to provide a visual technique in the analysis of textual information. It can be seen as a natural extension to textual relational databases and as such was quite successful in providing the user-interface which was both easy to use and effective in its retrieval of simple information from a relational database.

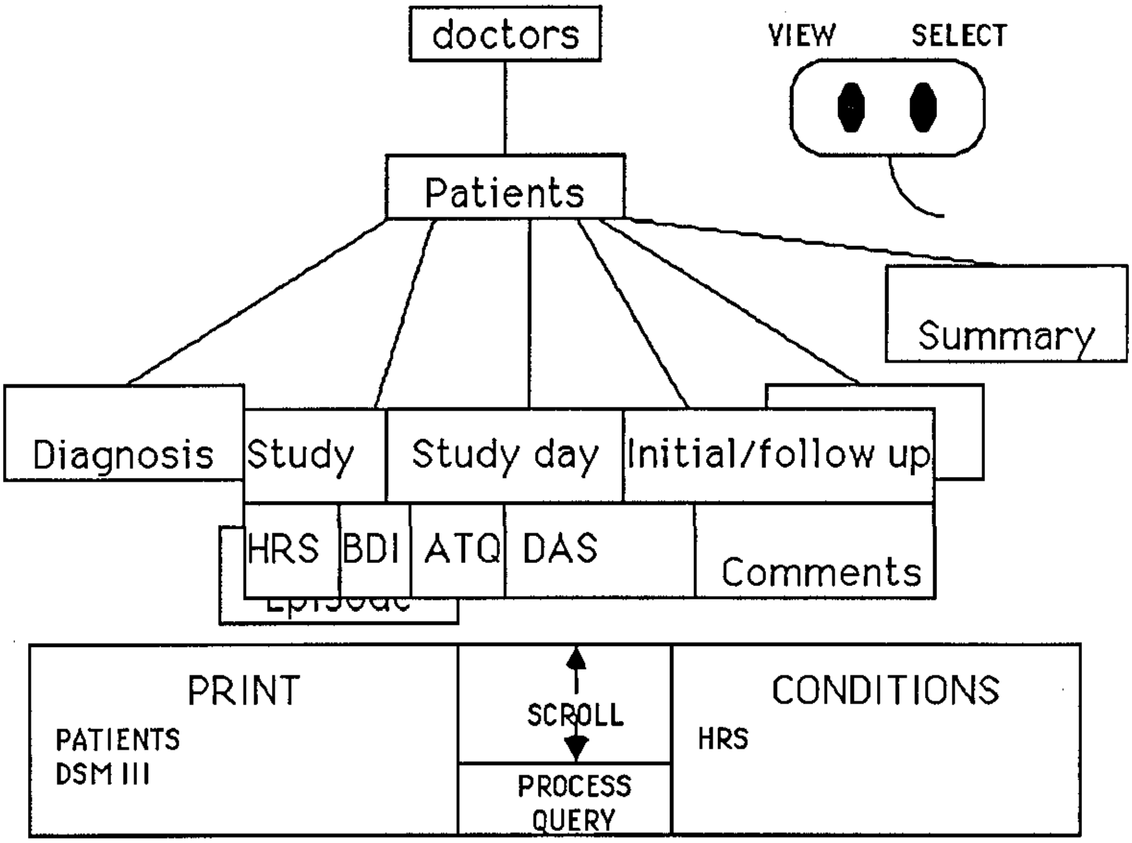
3.3.2 Texas Health Centre Graphical Interface

A more advanced example of a graphical database system was that developed for physicians at the University of Texas Health Science Centre at Dallas (Burgess, Leigh and Ali 1986). The purpose of this system was to enable physicians to quickly process their own database queries from information entered by nurses without spending time learning a complex query language. The interface shown in figure 3.3 provided the user with a graphical model of the structure of the database. The top level of the model gives a number of headings which correspond to the highest level of database information the physicians are interested in. Each heading appears in a box which has its relationship to other headings indicated by connecting lines.

Within the interface shown the lower portion of the display is split between two windows, PRINT and CONDITIONS. The print window holds the names of data items selected for printing while the conditions window specifies the items chosen. The 'Scroll' box enables the user to move the highlighted bar within a window, effectively scrolling a window up or down so as to view all of the items within that window. A mouse allows various different levels to be selected. This mouse is operated by moving the cursor over a heading box and pressing the 'view' button displaying a lower menu. Clicking the mouse upon the select button chooses an item from this lower menu. The menu chosen in the example shown below is the Tests menu. After selecting with the mouse this Tests menu expands to display information relating to the nature of the tests such as study and study day.

This system differs from Zloof's graphical database interface in that it has a more complex visual interface with the ability to scroll the screen, there is also the ability to view and select particular attributes from the textual database using a mouse. It provides a hierarchy of items to select from, and is visually more sophisticated than Zloof's example. The major function of this system is to provide a visual interface which can be easily used by inexperienced staff who have little knowledge of computers. This was a significant advance over the purely textual data encountered within earlier relational databases.

FIGURE 3.3



EXAMPLE OF USE OF AFFECTIVE DISORDERS UNIT DATABASE

3.4 IMAGE DATABASES

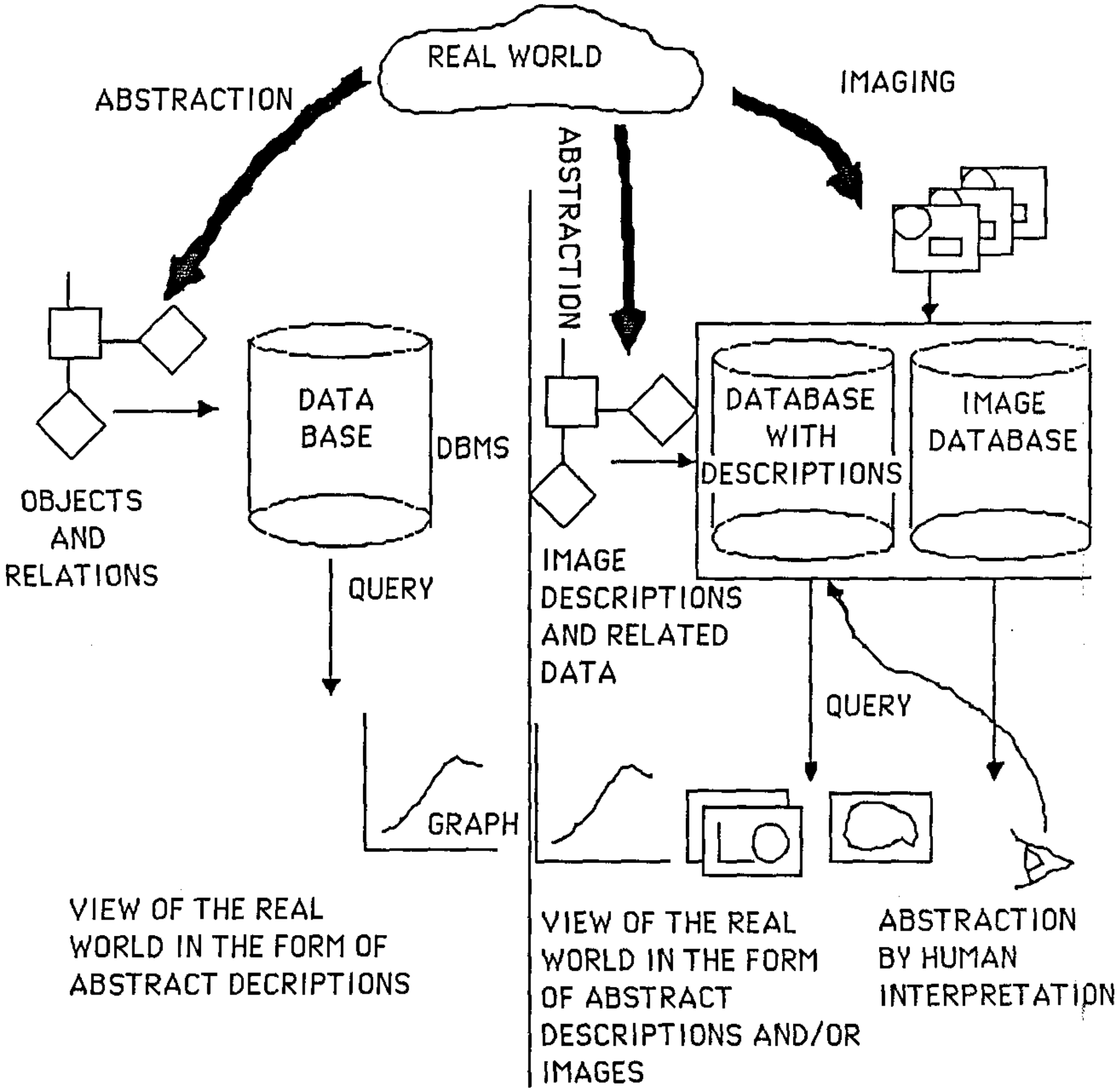
3.4.1 Image Database Introduction

Many database systems require that pictures as well as text are stored within the database. People in general prefer pictorial representations to words, the reasons for this include the fact that the speed at which pictures can be seen and interpreted is considerably faster than that possible with textual data (Potter and Faulconer 1975). Psychologically pictures offer a highly parallel mechanism for information transfer whereas text has a relatively slow rate of assimilation due to its serial nature (Barker and Manji 1989). Pictures also have many striking characteristics such as brightness, colour, contrast, strange arrangements of patterns, etc. Visual memory and recognition of pictures containing such characteristics is excellent (Noton and Stark 1971).

Pictorial databases of images differ from textual databases in that they visually display pictorial data rather than use textual descriptions of attributes. A picture is therefore much more specific in its representation than a corresponding textual description since different users may specify different textual descriptions of identical pictures. From the graphical database manipulation principles detailed earlier in this chapter, pictorial database management systems were developed in order to aid the user in the manipulation, storage, retrieval and analysis of pictorial information. These pictorial database management systems contained images, an image being a two-dimensional representation which has either been constructed, as in the case of a drawing, or captured, as in the case of a videodisk or digitised picture.

There are important differences between data stored in a conventional database and a database containing images. In the design process of a conventional database the part of the real world to be modelled is subject to interpretation and abstraction during its transformation into database objects. In general the level of abstraction is very high. However the images themselves are the objects in an image database and their level of abstraction is low. So the main difference is due to the fact that abstraction takes place on different levels. Owing to the low degree of image data abstraction, a variety of modes of storing, coding and displaying images is possible (Assmann, Venema and Hohne 1986).

FIGURE 3.4



THE FUNCTION OF A CONVENTIONAL DATABASE SYSTEM VS THE FUNCTION OF AN IMAGE DATABASE (Assmann, Venema and Hohne 1986)

3.4.2 What is an Image Database System?

An image database may be formally defined as:-

A system in which a large amount of image data and its related information are integratedly stored (Chang 1981). Using this definition a simple collection of images without any management system should not be regarded as an image database irrespective of the number of images present.

3.4.3 Image Database Characterisation

Tamura and Yokoya (1984) have suggested that image-like information generally belongs to one of the following three classes:-

1) Images represented by a matrix of elementary points and stored in this form as a compacted structure that retains all the image information. This kind of information is generally obtained via photographic-like processes.

2) Graphics, shapes or objects represented by their contours (points, vertices, vectors, segments) and by some other associated information (colour, origin, characteristics). This data may be obtained by digitisation, image processing or automatic generation with mathematical functions.

3) Systems in which the images are separated from databases describing the images.These databases contain descriptions about the nature and form of such information and are stored within a textual database. Such a system may be regarded as one overall database or as supporting separately stored databases.

Due to its very peculiar nature and large diversity, image information is not easy to manipulate, even for such classical database operations as retrieve or update. It is however not difficult to retrieve an image by its registrations such as name or number, but it is far more difficult to retrieve it by selection criteria that are contained pictorially within the image. However systems which are used for retrieval of pictures are very different in their design from systems that are used for the processing and manipulation of pictures. Retrieval oriented systems often require methods of access to the image data such as the analysis of textual information abstracted from the images, whereas processing systems usually require detailed digital image processing techniques when analysing images.

3.4.4 Examples of Practical Image Data Bases

Image-oriented applications, and more generally those applications dealing with image-like information, have become both increasingly numerous and more diverse in nature. Two of the most popular forms of such image databases are image catalogues where a videodisk takes the place of paper as a medium for recording pictures, and secondly, guidance systems (civil or military) where the objective is to display maps and similar data enabling an individual to navigate around a physical location (such as a town) without ever having been there. However the most common application areas for image databases can be broadly subdivided into three main headings:- medical imagery, computer-aided design (CAD) and geographical information processing.

1) Medical Imagery

Image databases are commonly used in the medical domain but most of the existing systems were designed to handle image registrations about patients rather than attempting any detailed analysis of digitised images such as X-ray results. Image databases such as ISQL (Assmann, Venema and Hohne 1986) select images based upon name of patient and the date at which a corresponding photograph was taken rather than information directly extracted from images. The problems here are twofold: not only is there an enormous amount of data to analyse, but also medical and practical questions arise in taking account of this data.

2) Computer Aided Design (CAD)

CAD is a very specific kind of application whose needs are well defined. These systems store spatial data as graphic information and are designed for draughting and object design/modelling (Exler 1988). Problems in this area are mainly due to the specific character of CAD objects, the structure of which may be very complicated. Image databases within CAD usually require different views of objects and powerful image manipulation techniques upon which selection criteria can be based (Spooner 1986).

3) Geographic Information Processing (GIP)

Geographic oriented applications contain such diverse domains as urban and regional planning, weather forecasting, protection of the environment, the use of natural resources, or simple cartography. Such systems were amongst the first of the image-oriented applications to make use of the storage and processing facilities of computers. This has lead to the development of geographical information systems containing a vast quantity of geographical data. Recently a requirement for handling both aerial and satellite imagery data has arisen (Bylinsky 1989) and since this data is by its very nature different to more conventional images (for example different image resolutions are required over cities and open land) then specific data structures are needed in order to take account of such data.

Rather than discussing in depth all three of these application areas the development of geographical information systems will be examined. One of the most important reasons for concentrating on geographical information systems is that these are one of the few application areas which have made attempts at basic spatial analysis of objects within images. Advances within other application areas has broadly mirrored those within these systems, many of the same principles used within geographical information systems (such as storage using relational databases) being applied to other application areas.

3.5 GEOGRAPHICAL INFORMATION SYSTEMS

A Geographical Information System (GIS) is:-
A decision support system involving the integration of spatially referenced data in a problem solving environment (Cowen 1988). Geographical data is data about specific locations on the earth and provides information such as spatial data, patterns of housing, land use, voting and health information. This information is useful to many sections of the community including local government and the armed forces. However a GIS's system's purpose is first and foremost spatial analysis.

Geographic databases currently in existence are severely inefficient, in terms of both compaction and speed of use, as well as rigidity and narrowness in the range of applications and data types which can be supported by a single database. These efficiency, versatility and integration problems can be traced in large part to the profound differences in the storage formats commonly used for spatial data handling. The basic problem is a lack of understanding of the nature of spatial data, and a lack of a unified body of knowledge of the design and evaluation of spatial data structures (Peuquet 1984).

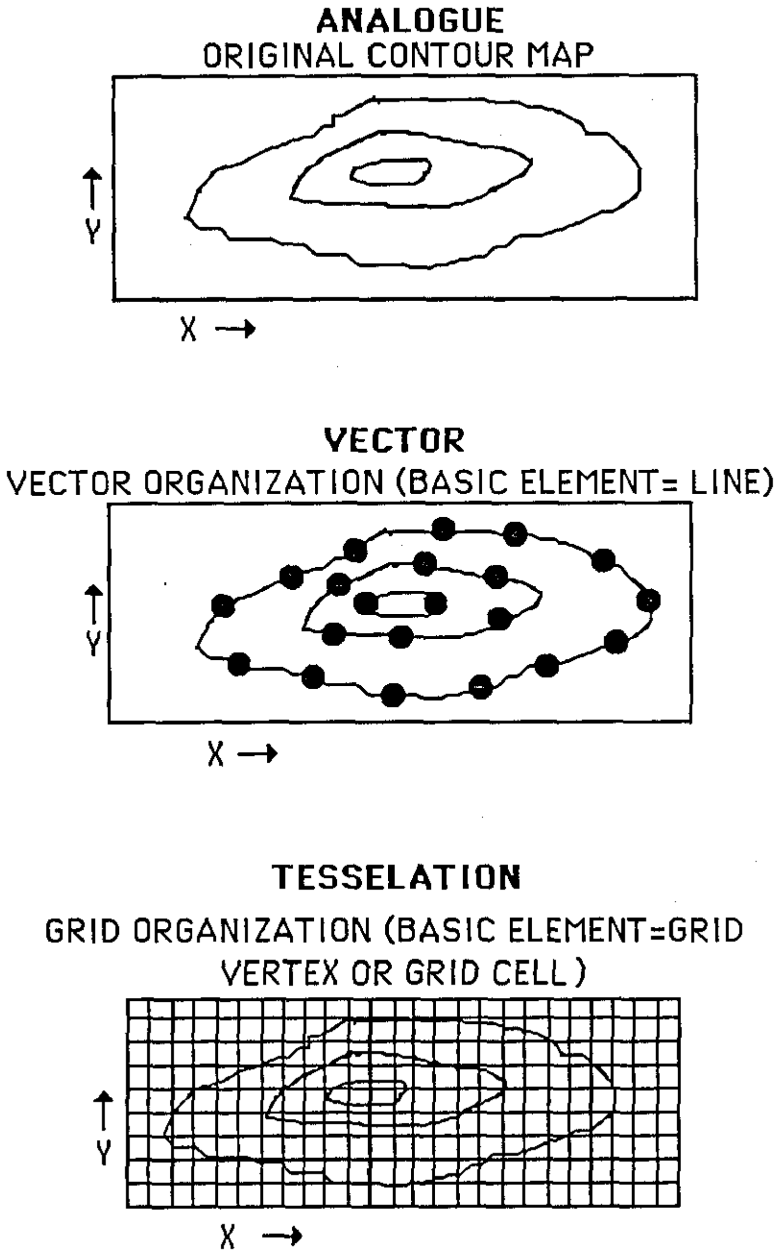
Conventional data models are not particularly suitable for geographic applications because they do not efficiently support the type of operations that are required for geographical analysis, nor are they suitable for the storage and manipulation of spatial and graphical data (Ooi, Sacks-Davis and McDonell 1989). In general when geographical pictorial information is represented, this is achieved in one of two ways, either as a vector (topological) model or as a non-vector type model (Chang and Kunii 1981). However non-vector spatial data models encompass much more than data models based on a raster square mesh. This class includes any infinitely repeatable pattern of a regular polygon or polyhedron. The term in geometry for this is a 'regular tessellation'. Examples of these two forms are shown in figure 3.5.

3.5.1 Vector (Topological) Structures

Vector data structures are a family of ordered point sets, including isolated points, line segments or point pairs. In the simplest systems the data is stored by lists of points outlining geographical features with no explicit relationships among the features (Chang 1981). For example the corners of a rectangular field may be recorded by manually moving a cursor to each of the four corner points and recording their coordinates. If the sides are straight then this information is sufficient to describe the outline of the feature. An extension of this method is to chain all line segments between branches or intersections to form point lists, starting with a "from" node ending at a "to" node.

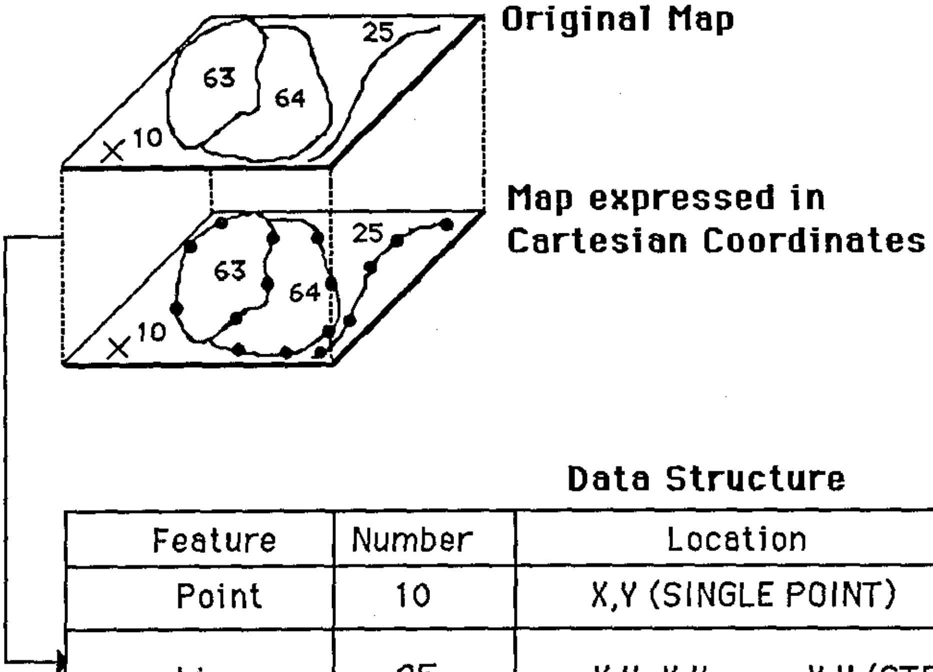
The simplest vector data model for geographic data is a direct line-for-line translation of the paper map. Here each entity on the map becomes one logical record in the digital file, and is defined as strings of x-y coordinates. This structure is very simple to understand since, in essence, the map remains the conceptual model and the x-y coordinate file is more precisely a data structure. The two-dimensional map model is therefore translated into a list, or one-dimensional model. A polygon recorded in this manner is represented by a closed string of x-y coordinates which define its boundary. For adjacent polygon data, this results in recording the x-y coordinates of shared boundary segments twice - once for each polygon. Although all entities are spatially defined, no spatial relationships are recorded. A digital cartographic data file constructed in this manner is commonly called a 'spaghetti file'. An example of this spaghetti model is shown in figure 3.6.

FIGURE 3.5



BASIC TYPES OF SPATIAL DATA MODELS

FIGURE 3.6



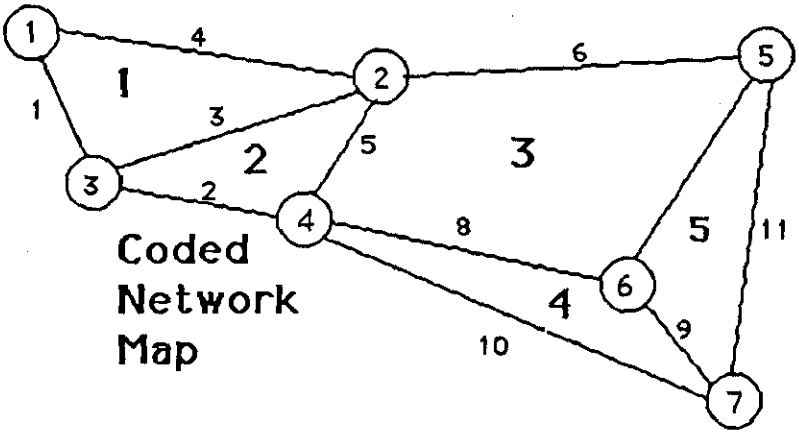
Data Structure

Feature	Number	Location
Point	10	X,Y (SINGLE POINT)
Line	25	X_1Y_1,X_2Y_2..........X_nY_n(STRING)
Polygon	63	X_1Y_1,X_2Y_2..........X_1Y_1(CLOSED LOOP)
	64	X_1Y_1,X_2Y_2..........X_1Y_1(DATA STRUCTURE)

THE 'SPAGHETTI' DATA MODEL

However the most popular method of retaining spatial relationships among entities is to explicitly record adjacency information in what is known as a topologic data model. Here the basic logical entity is a straight line segment. Each individual line segment is recorded with the coordinates of its two endpoints. In addition, the identifier, or name of the polygons on either side of the line is recorded. In this way the more elementary spatial relationships are explicitly retained and can be used for analysis. In addition this topological information allows the spatial definitions of points, lines and polygon-type entities to be stored in a non-redundant manner. A simplified example is shown in figure 3.7 (Peuquet 1984).

FIGURE 3.7



Link number	Right Polygon	Left Polygon	Node 1	Node 2
1	1	0	3	1
2	2	0	4	3
3	2	1	3	2
4	1	0	1	2
5	3	2	4	2
6	3	0	2	5
7	5	3	5	6
8	4	3	6	4
9	5	4	7	6
10	4	0	7	4
11	0	5	5	7

Topologically Coded Network & Polygon File

Node Number	X Coordinate	Y Coordinate
1	23	8
2	17	17
3	29	15
4	26	21
5	8	26
6	22	30
7	24	38

X,Y Coordinate Node File

THE TOPOLOGICAL DATA MODEL

The management and retrieval of geographic information is more flexible when data is stored in vector format. Individual features, such as boundaries and polygons are specifically coded and therefore individual features can be more easily manipulated or searched for. However data has to be encoded which is itself time consuming.

Information about vector-based data is often stored in textual databases enabling textual query languages to be used to find information within them. However the complexity of this data, its large volume and high resolution display and printing requirements placed large demands on the database management systems used. These vector based techniques were one of the initial attempts at the analysis of geographical information, however they provided only the barest minimum of information concerning the nature of the objects contained within the images. Such techniques have been primarily used for mapping boundaries, rivers and roads since these objects do not require detailed physical descriptions. Such techniques differ from raster formats in that encoding of the objects and shapes within images are specified. Therefore vector formatting uses an abstraction of the geographical information from an image in order to represent data.

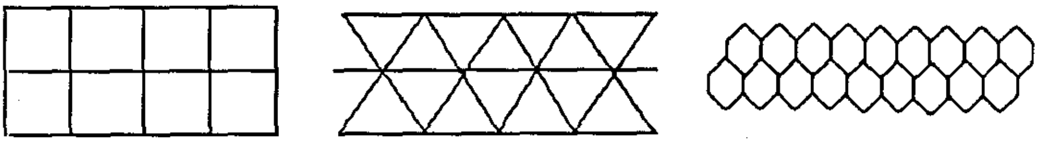
3.5.2 Raster and Tessellation Formats

A raster format consists of pixels in a rectangular matrix, or a raster of dots that cover an entire screen (Newman and Sproull 1979). Algorithms and data structures have been developed which enable raster formatted information to be manipulated therby encouraging the corresponding development of pictorial database systems. These raster formatted information systems have tended to evolve from image analysis software developed for remote sensing applications. The resulting information is held in store in an array or matrix whose lines and columns correspond to a grid. However data in raster format occupies considerably more storage space than vector formatted data. Techniques exist for reducing the amount of data stored, where for example there are large blank areas on a map, but inevitably the storage requirements for raster map data are much greater than for vector coordinates.

The most important aspect of raster formatted information is the fact that the digitised information can be directly analysed using image processing techniques, therefore the effectiveness of a pictorial database using digitised data depends to a large extent on the resolution of the instrument taking the picture and the ability of the image processor to detect objects within the digitised image produced.

Unfortunately image processing techniques are at present limited in their capabilities, this is because the algorithms which enable pattern recognition of objects are not yet sufficiently advanced such that objects (unless they are very simple) can be recognised. However raster formatting does allow analysis of the actual image rather than a study of an abstraction of the data, as in the vector case. Tessellation models can be seen as an extension to raster models (Peuquet 1984). In general there are three forms of regular tessellation, square, triangular and hexagonal meshes, of which square models (using a raster format) can be seen as the simplest. These three models are shown in figure 3.8.

FIGURE 3.8

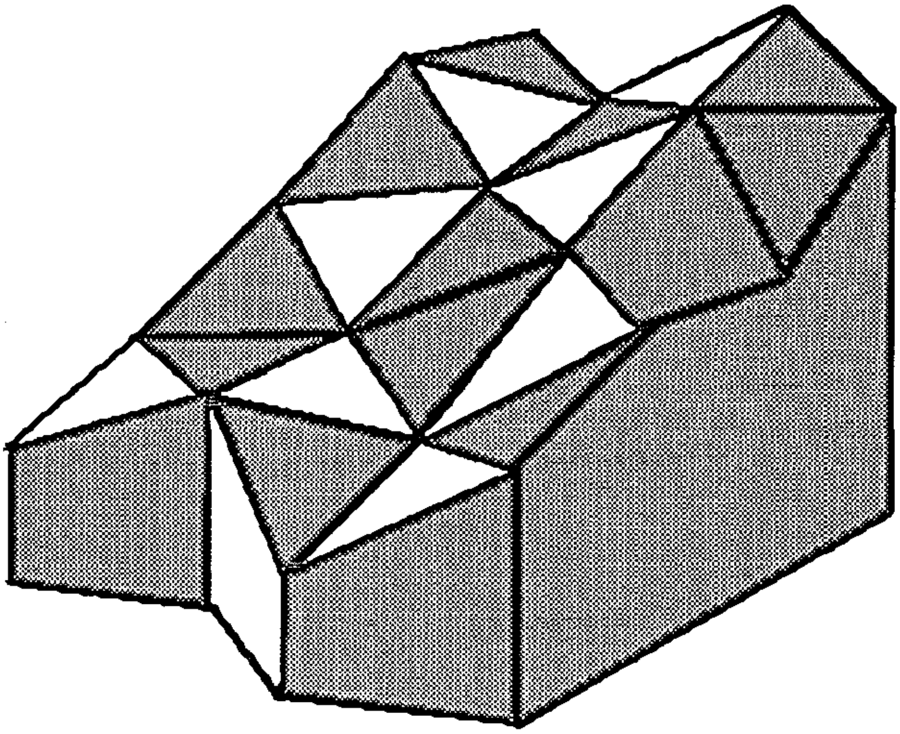


THE THREE REGULAR TESSELLATION MODELS.

The primary advantage of the regular hexagonal mesh is that all neighbouring cells of a given cell are equidistant from that cell's centre point. Radial symmetry makes this model advantageous for radial search and retrieval functions.

Regular triangular meshes, are rarely used for the representation of geographical data, irregular meshes often being used instead. A contributing factor in the almost total lack of use of the regular triangular mesh for surface data is simply that such data are normally not captured in a regular spatial sampling pattern (Ahuja 1983). An irregular mesh representing geographical data is shown in figure 3.9.

FIGURE 3.9



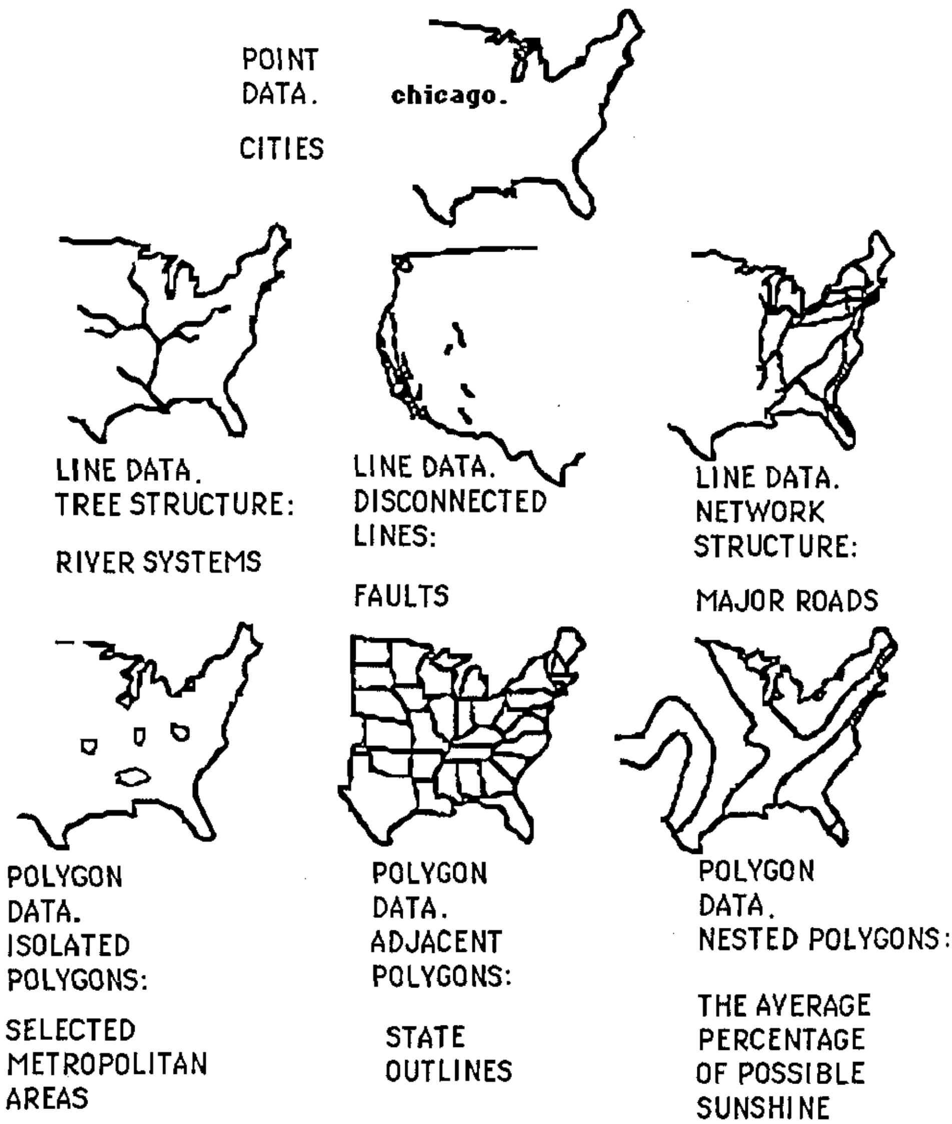
A IRREGULAR TRIANGULATED NETWORK REPRESENTING SURFACE DATA

3.5.3 Data Structures for Geographical Image Data

Geographical databases currently in existence are inefficient, in terms of both compaction and speed of use, as well as rigidity and narrowness in the range of applications and data types which can be supported by a single database. These efficiency, versatility and integration problems can be traced in large part to the profound differences in the storage formats commonly used for spatial data handling. The basic problem, however is a lack of understanding of the nature of spatial data, and a lack of a unified body of knowledge of the design and evaluation of spatial data structures (Peuquet 1984).

Two central issues are common to the design problems encountered in image databases. Firstly how do you provide a unified approach to retrieve and manipulate pictorial information, and secondly how do you utilise data structures to improve and develop algorithms for pictorial information retrieval and manipulation which possess good spatial properties? In general there are three different kinds of spatial data in geographic pictures, these are point features, line features and polygon (or area) features (Peuquet 1984). Examples of these three features contained within a map of the United States of America are shown in figure 3.10.

FIGURE 3.10



EXAMPLES OF SPATIAL DATA TYPES FOR A MAP OF AMERICA

Point features may be small areas or junctions of roads or a position indicated by its coordinates. Line features are features such as roads in a road map or rivers and railways, which can be considered to consist of simple lines. These line features are represented by several x-y coordinates. Other attributes may be associated with these line features, for example attributes for a road might be its material, width and curvature.

Polygon feature examples include lakes airfields and parks and may be represented by coordinates of a polygon or by groups of pixels. On a given map these features may *intersect*, may be *adjacent* to each other or may *contain* other object's. Their relationships are therefore termed spatial relationships and are object-oriented rather than attribute based (Ooi, Davies and McDonell 1989).

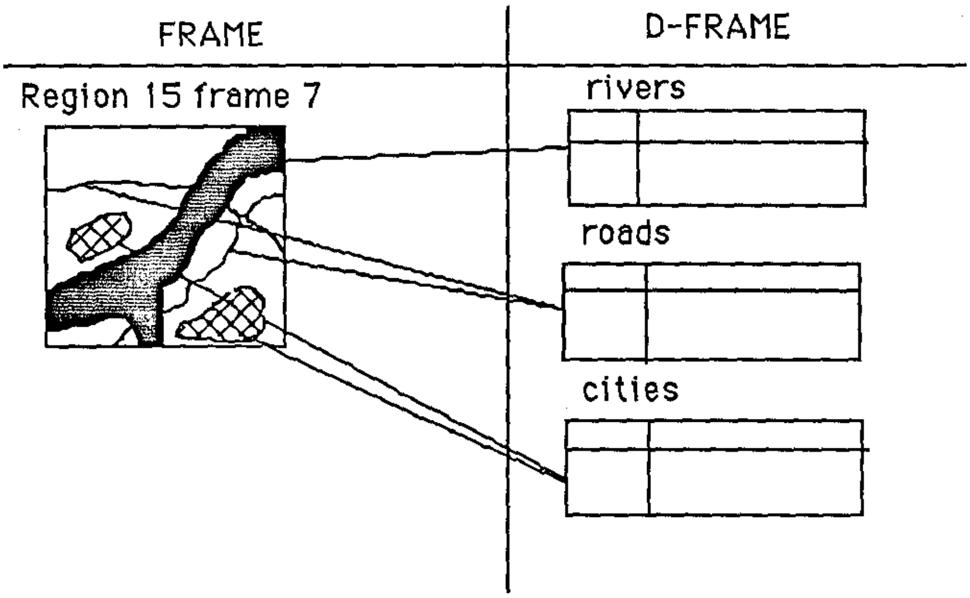
The three main data structures for a traditional textual database are the relational, the hierarchical and the network models, two of these the relational and the hierarchical can be modified and extended into the pictorial domain.

3.5.3.1 <u>The Relational Model</u>

Pictorial information may in general include geometrical descriptions of an objects shape, the attribute description of the object and the spatial relationship between objects. Therefore there is a need for a generalised information system which can manipulate both tabular data and graphical data as well as image data. The relational model presents a relational database approach for describing complex pictures, the tabular structures it employs are conceptually not difficult to understand and are easy to access. For these images a picture object may be defined as part of a picture that can be operated on by textual picture query languages. Picture objects may be simple objects such as rivers, lakes and airfields or they may be composite objects with varying degrees of detail.

In applying the relational method images are subdivided into frames. Under such circumstances the image is called the map and the relational pictorial database storing the abstracted data is called the d-map (Chang, Lin and Walser 1979). The map is composed of a collection of picture overlays representing different geographical features which may be defined statically or created dynamically through the manipulation of picture operations. It is only when all these picture overlays are combined together (for example a city overlay combined with river overlays) that the complete map is obtained. Therefore a map is not stored in the relational database in the way it appears in the window to the user, rather these maps are generated by processes that transform relational data into a visual form. The process of converting relational data contained within a d-map into a map is called materialisation. The inverse, converting a map into a d-map is called abstraction. A frame is the smallest unit for visual display and its logical representation within the relational database is called a d-frame. A d-frame is a set of relations from which a single frame may be loaded. Each relation in a d-frame corresponds to features of the same class. The decomposition of a frame into a d-frame is shown in figure 3.11.

FIGURE 3.11



EACH FRAME CAN BE DECOMPOSED INTO ITS COMPOSITE ATTRIBUTES

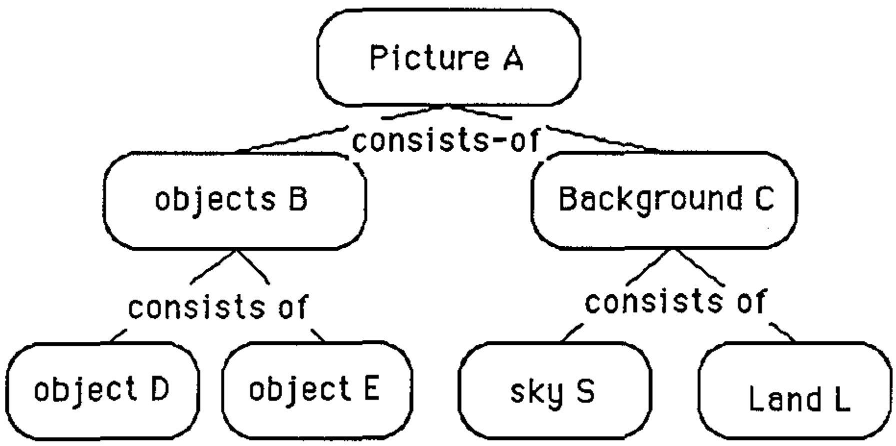
Pictorial query languages are used to query the relational data stored about the images and their component overlays. However these languages rely upon a good understanding of relational databases and are often quite complex. They are also severely limited in the questions which they can ask about the pictorial information. Spatial data concerning the relationships between objects present are particularly difficult to extract.

An integrated pictorial data-base management system was developed by Chang, Reuss and McCormick (1977) that combined a relational data-base management system with an image manipulation system enabling the user to perform various pictorial information retrieval functions by zooming and panning in order to manipulate the pictorial database. An example of such an image is a picture of a town in which the image may initially be complex, but with logical zooming the image may be decomposed into simpler overlay elements such as roads, parks, rivers and railways. This concept of overlays and frames can be applied to a large number of images some of which can be very complex.

3.5.3.2 Hierarchical Systems

An alternative to the relational database approach are hierarchical systems. These consist of an image information system in which an object is described in terms of its shape, symbolic description, and relations with other objects (Sties, Sanyal and Leist 1976). Objects in a picture can be retrieved in terms of symbolic descriptions and/or object relationships.

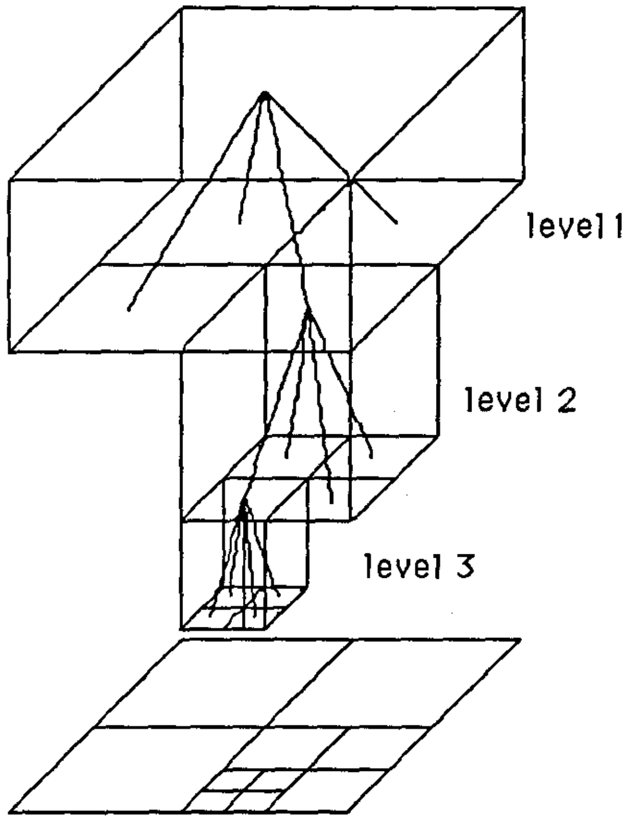
FIGURE 3.12



OBJECT RELATIONSHIPS WITHIN AN IMAGE

A alternative hierarchical data structure scheme can be used to store images by the regular decomposition of pictures into adjacent quadrants with different resolution. Each quadrant corresponds to a node in a tree, called a quad-tree. Images are stored hierarchically with different resolutions. A representation of the database structure consisting of these resolutions is shown in figure 3.13.

FIGURE 3.13



HIERARCHICAL DATA STRUCTURE FOR IMAGES

3.5.3.3 Complex Spatial Data Structures

As discussed earlier, the vector format is used to store geometrical data in the form of points, lines and polygons. Additional structures are needed to store the relationships between these objects (topological data). In the literature several solutions have been proposed for this multi-dimensional data such as k-d trees, quadtrees, Peano keys and grid files (Oosterom 1988).

All the models so far discussed both relational and hierarchical have attempted to describe pictorial information in terms of textual abstractions. These textual abstractions are then incorporated into the system allowing the user to query the pictorial database. A more detailed description of some of these textual picture query languages now follows.

3.6 PICTURE QUERY LANGUAGES FOR DATABASE SYSTEMS

A picture query language can be defined as a textual or pictorial language which can be used to analyse data relationships within a pictorial database (Chang and Fu 1981). Important information about images include their resolution, size, the time at which the images were taken, their storage and location.

However it is difficult for conventional textual query languages to express some types of query. Requirements such as the updating of picture descriptions and the retrieval of images through specified picture descriptions, require both manipulation of descriptions of the images as well as the manipulation of the images themselves. Picture query languages have therefore been developed in order to cope with such problems. These picture query languages are usually designed as additions to conventional query languages and contain a picture querying facility as well as the manipulation capabilities of conventional query languages.

3.6.1 Examples of Pictorial Query Languages

3.6.1.1 IMAID

IMAID (Chang and Fu 1980) is a relational database system interfaced with an image analysis system. By using pattern recognition and image processing manipulation functions, pictorial descriptions can be extracted from images. These extracted descriptions and sections of pictures are then integrated into a relational database, the original two-dimensional pictures are kept in a separate picture store. Thus queries about pictures can be manipulated through the relational database, eliminating the need to process vast amount of image data. If the user's request can be expressed in terms of the extracted picture descriptions, there is no need to retrieve and process the original pictures. If on the other hand the extracted information is not sufficient, all pictures satisfying selection criteria can be retrieved from picture store and processed until the required precision is obtained.

A processing package is constructed and stored in IMAID for the processing of each image type. A processing package consists of several processing sets, which extract a desired description of the specified image type. Each set consists of elementary processing functions used in pattern recognition and image processing. These include texture measurement, filtering, edge detection, segmentation, primitive extraction and selection, syntax analysis, and classification functions. For example in the analysis of Landsat images, the processing sets ROAD, RIVER, COMMERCIAL-AREA and MEADOW are used. In the case of the ROAD processing set the desired pattern is a set of connected pixels that satisfy the ROAD grammar. The road network of an image can then be obtained as a line drawing after execution of the ROAD processing set.

3.6.1.2 QPE (Query by Pictorial Example)

Query by picture example (QPE) is the picture query language method used for IMAID.

The manipulation capabilities of QPE fall into six fundamental types.

1) Conventional manipulation:

Conventional manipulation allows the user those relational database queries permitted within textual relational databases.

2) Pictorial entity construction:

A desired pictorial entity can be constructed as a point, a line segment or as a region from the intersection, union, or negation of the originally stored or previously constructed pictorial entities abstracted from the digitised map. However a line segment is considered to be the basic unit in representing the IMAID pictorial entities. This is because a line is represented by a set of line segments; a point is a line segment with two coincident endpoints and a region is represented by its boundary, a closed line. These entities can be used to find sections of lines that lie within areas, such as roads that lie within city limits, giving the required information in the form of the intersection points between the road and the boundary enclosing the city (Chang and Fu 1981).

3) Pictorial attribute manipulation:

Features such as the area of a region, the length of a line segment, the distance between two given points or the perimeter of a region can be obtained by examining abstracted data. For each pictorial operator, the system provides a procedure that performs the required operations with supplied parameters. These operators are intended to construct only the basic relations among points, lines and regions for queries concerning pictorial information.

4) Image-sketch-relation conversion:

Pattern recognition and image processing techniques allow structures and features of the original raster image to be extracted via analysis of a representation of the image called the picture sketch. Therefore applying processing sets allows data to be extracted from images and stored within the relational database.

5) Pictorial example:

In addition to textual formulation of queries, selected portions of a displayed picture can be used to specify further queries. These queries can be shown by having the desired positions pointed out by the user on the display terminal, an example being pointing to a road in order to determine its name.

6) Similarity manipulation:

Entities similar to a given entity can be retrieved. Here image frames most similar to the images shown on a display terminal can be located. This similarity of images is based upon to the similarity of relational database information about items within them such as the similarity of road networks.

IMAID's effectiveness depends to a large extent on the ability of the pattern recognition and image processing functions to identify pictorial information within an image. If this information is not stored in the relational database then attempting to find the correct information is much more difficult. Having processed such data into a relational format, IMAID then relies upon textual similarity between objects in the relational database to find matching objects within the image database. Although the system allows objects (such as cities) to be pointed to in order to identify their names, it has no direct means by which groups of objects can be effectively located from the images based upon the two-dimensional spatial arrangement of the objects within an image.

3.6.1.3 <u>Other Picture Query Languages</u>

Graphics-oriented Relational Algebraic INterpreter(GRAIN) is an extension of RAIN the relational algebraic interpreter (Chang, Reuss and McCormick 1978). GRAIN is characterised by the introduction of the concept of logical pictures which can be considered as the result of extracting meaningful parts from the entire image. Logical pictures are defined in three relational tables called the

Picture Object Table (POT), the Picture Contour Table (PCT) and the Picture Page Table (PPT). The Picture Object Table describes the attributes of picture objects and the relationship between picture objects. The Picture Contour Table is used to define picture objects using contour codes. The Picture Page Table relates the mapping from logical pictures to physical pictures. One of the characteristics of this picture query language is that it can translate user queries specified in GRAIN syntax language into picture algebraic operations.

IQ (Lien and Harris 1980) is another picture query language but unlike GRAIN and IMAID, IQ stores only image sections and the images themselves. The user interacts with the data with the aid of the IQ query language which is designed to support image editing, transformation, storage retrieval and display. The system maintains five kinds of files, image, window, transform, colour and zoom.

Within all these picture query languages users are required to learn complicated textual notations in order to obtain information from an image database. Textual picture query languages only have a limited degree of success in retrieving pictorial information from a pictorial database because of the fundamental differences between pictures and text. This is because in many cases the functions of the query language do not meet the user's requirements. Thus when searching for pictorial scenes using a textual picture query language users are unable to express their query because of the restrictions of the language supplied. Such pictorial query languages are also very restricted in the domain that they can be applied to, each language is usually designed for one application area and has operations which can only be applied to the attributes and constituents of that type of image. For example in the case of geographical applications textual pictorial query languages are designed to analyse and extract entities such as roads, rivers and other geographical objects.

We also suggest that pictorial query languages require an understanding of concepts such as relational databases and the ability to specify only that information required from an often wide range of attributes. These queries need to be expressed in the form of mathematical notations such as logical operators and predicate calculus, further confusing many users who wish only to find specific objects or particular spatial arrangements of objects within an image. Thus this complexity ensures that often only experienced users with a thorough understanding of such concepts can gain any real benefit from the use of such languages. However the major disadvantage of these textual picture query languages is that such languages do not allow users to easily analyse the spatial arrangements of the objects contained within images . Therefore users are limited in the way that they can analyse the information given. It is suggested that the ability to use a more visual technique from which spatial information can be both selected and analysed would help users querying pictorial databases. This will be further discussed in later chapters.

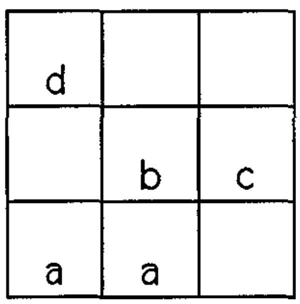
3.7 MODERN TECHNIQUES OF TEXTUAL SPATIAL ANALYSIS

Two modern methods of managing pictorial information which attempt more detailed analysis of spatial information within pictorial databases will now be discussed.

3.7.1 An Intelligent Image Database System

Perhaps the most important problem for the design of image databases is how the images should be stored. For more intelligent and faster systems the data structure should be object-oriented and should capture as much spatial knowledge as possible (Chang, Yan, Dimitroff and Arndt 1988). A new technique recently suggested is the representation of a picture is by a two-dimensional string .An example is shown in figure 3.14

FIGURE 3.14



REPRESENTATION OF A PICTURE f

Here a,b,c,d represent objects and the two-dimensional string representing the picture is

(a=d<a=b<c, a=a<b=c<d)

"<" denotes the left-right spatial relation
"="denotes the spatial relation "at the same spatial location as"

Therefore this two-dimensional string representation can be seen to be the symbolic projection of picture f along the x and y directions.

The symbolic picture is the mapping

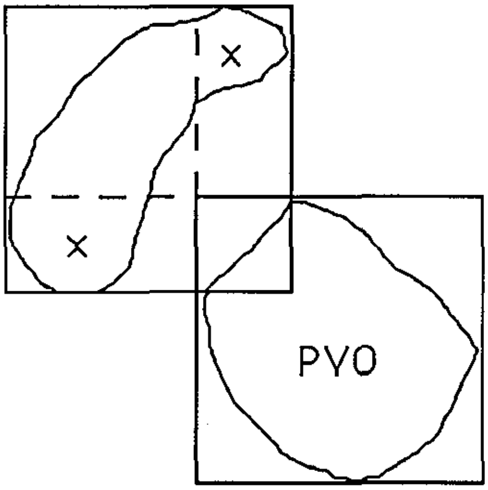
f(1,1)={a} f(1,2)={ } f(1,3)={d}
f(2,1)={a} f(2,2)={b} f(2,3)={ }
f(3,1)={ } f(3,2)={ c} f(3,3)={ }

From a given f the corresponding two-dimensional representation can be constructed and vice versa. A query can be specified graphically by drawing an iconic figure on the screen of a computer, this can then be translated into the two-dimensional string representation. These iconic indexes can also be used for browsing. Here the spatial relations among objects are called orthogonal relationships wherein the relations are defined in terms of their minimum enclosing rectangles (MER). Using this method three types of spatial relations can be identified. These are:-

1) non overlapping MER
2) partially overlapping MER
3) completely overlapping MER

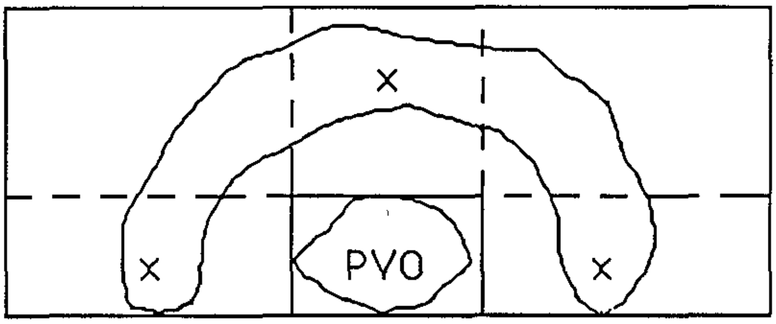
The idea here is to regard one of the objects as a 'Point of View' Object (PVO) and then view the other objects in four different directions (north, south east and west). Two examples are

FIGURE 3.15



THE POINT OF VIEW OBJECT AND ITS CORRESPONDING ORTHOGONAL RELATIONS FOR PARTIALLY OVERLAPPING MER

FIGURE 3.16



THE POINT OF VIEW OBJECT AND ITS CORRESPONDING ORTHOGONAL RELATIONS FOR COMPLETELY OVERLAPPING MER

The sub-objects segmented by the 'point of view' object are called ortho-relational objects of the original object and are marked with an x in the example. Therefore by encoding the values of the x's the two-dimensional string representation of the shape can be obtained and from this the symbolic picture.

FIGURE 3.17

THE ORIGINAL AND SEGMENTED IMAGE

The two-dimensional string encoding of the segmented image in the above case is therefore

c1<c2 A <B c4<c3, B<c2<c4<c1 A c3

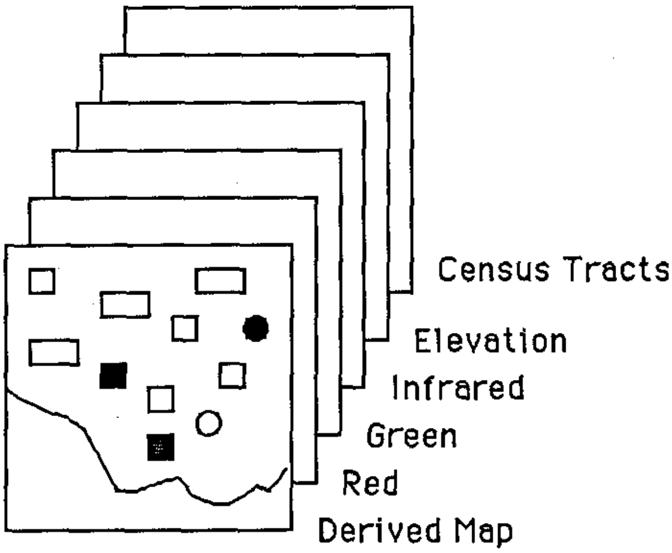
From this information the symbolic picture can be built and manipulation of this information can be achieved.

Although this system goes some way to taking account of the spatial layout of the information stored within the image database, the means by which the segmented image is stored as a two-dimensional string is somewhat cumbersome and long winded for the case of a complex picture. Again as with earlier examples the process of querying the image database is done by textually querying information extracted from the image database and stored in a relational table, rather than being actually able to pictorially query the images using visual techniques. However this system does at least go some way to taking account of the spatial layout of geographical information albeit in a somewhat mathematical way.

3.7.2 PICQUERY A High Level Query Language for Pictorial Database Management

PICQUERY (Joseph and Cardenas 1988) has been designed to reside above a Pictorial Database Management System (PICDMS) and be able to operate on more general areas than specific application areas such as geographical data. PICQUERY is intended to be the interface through which the user may access conventional relational databases using Query By Example (QBE). To the user both PICQUERY and QBE are seen as one single language. The PICDMS uses a grid data representation scheme and has an unique dynamic stacked logical data structure which consists of two-dimensional pictures on the same grid.

FIGURE 3.18



THE DERIVED MAP CONSISTS OF ALL THE OTHER IMAGES COMBINED

The PICQUERY language commands may operate on a whole pictorial database or a set of picture-object identifiers. Objects are identified to PICQUERY with the aid of a feature extraction program. Some PICQUERY commands operate on a group of objects in a picture while other operations apply to the entire database.

In the PICQUERY example below the user types in 'DISTANCE' and the system responds with the following table in which the user enters information into the appropriate columns (in this case MAINE into field OBJECT 1 NAME and NEVADA into the field OBJECT 2 NAME columns).

FIGURE 3.19

PICTURE	OBJECT 1 NAME	OBJECT 2 NAME	DISTANCE
PIC	MAINE	NEVADA	P

DISTANCE PICQUERY EXAMPLE

The minimum distance between the two objects is then calculated by the system and placed in the DISTANCE column.

Many other pictorial Manipulation operations can be carried out such as Horizontal Zoom, Rotating, Colour Transformations and Line Similarities. As before this example uses purely textual entry techniques by entering information into tables as a means from which to extract data from the image database. The effectiveness of PICQUERY depends greatly on the functionality of the feature extractor and its ability to identify individual objects. This is made easier by the fact that images are separated and only when these are combined is the full derived map produced. However there is still no facility to take account of the spatial layout of the objects contained with the various images that combine to produce the derived map.

3.8 CONCLUSIONS

Databases were initially designed for purely textual data. User-interfaces have been developed which employ graphical methods of querying information in an attempt to improve the ability of users to locate the information that they require These graphical database interfaces contribute to rapid understanding and accurate user performance. It has been further suggested that such systems also often allow the user to perceive the graphical interactions as manual table manipulations (Zloof 1975).

Pictorial database management systems have been developed using methods that modify simple textual database management structures. Initially the majority of pictorial database management systems used one of two structures. The simplest method was to use a simple text description for each picture in the database. A more formal technique was a fixed format header for each picture, this information had the advantage that it could be standardised. Although fixed format header systems attached textual information to a pictorial database in the form of registrations they did not attempt to extract visual information from the actual picture, such as values of attributes of the objects contained within them.

One example of a fixed format header system is the Image Structure Query Language (ISQL), a medical image database in which the image registrations consist of names of hospital patients, their age and date of the image showing their medical complaint (Assmann, Venema and Hohne 1986).

The next stage in the management of pictorial information was the use of relational database methods in which tables were constructed giving the relationships both between pictures and between objects within pictures after which the pictures where then textually queried. The two most common techniques here were the nominal and ordinal systems, nominal systems using the names of the pictures and objects within these pictures, and ordinal systems using the actual geographical coordinate values. More advanced techniques relied upon symbolic representations in which the pictorial information were broken up into its composite subimages until the detail required was achieved. These subimages were then connected by relationships in an attempt to provide a complete description of the pictorial information. Textual pictorial query languages were again used in the analysis of such relationships.

However as the number of pictorial databases have become more numerous, so more sophisticated textual methods have been developed in an attempt to describe the pictorial nature of the information within the images. This has led to a wide variety of textual retrieval techniques none of which has yet become the recognised standard for all applications. It is suggested that many users often find pictorial information contained within such systems difficult to query in terms of textual descriptions. This is because many of the aspects of a picture such as its appearance, its objects colours, shapes, sizes and orientations are remembered as visual images rather than as textual descriptions. Textual query languages therefore require a user to extract information from these visually memorised images in order to construct a textual query. It is further suggested that conventional textual query languages used within pictorial database management systems appear to be fundamentally designed for storage, retrieval and manipulation of alphanumeric data rather than as languages for the identification of information from images.

Thus our conclusion is that the major failing of pictorial database management systems discussed in this chapter is their inability to take account of visual techniques in the accessing and analysis of pictorial information. In many cases users of such pictorial database systems are often able to visually recall certain 'landmark' characteristics of pictures such as objects having a unusual shape, but are unable to readily retrieve this information from the pictorial database (Allport 1989). This is because although the landmarks are easily visually recalled, the methods of access to such data rely upon complex textual query languages along with knowledge of the names or coordinates of the objects. Queries from a geographical database such as "all the pictures in which roads, railways and lakes visually form a specific pattern" cannot be easily handled with a textual query language because the user is usually forced to use simple textual descriptions of objects. Information which relies upon location rather than name is especially difficult to handle textually, in that the spatial relationships between objects are often difficult to describe. The human brain is able to arrange things spatially very well and also remember the arrangement of these spatially located objects provided that they are visually observed (see later). It is suggested that using interactive systems in which the query is pictorially represented should be a useful tool in enabling the user to locate and identify information required and that textual queries of the same information (often in the form of semantic networks) are much less effective as a querying technique, especially where there is a large number of spatial relationships between objects within an picture. In the next chapter databases which have attempted to make use of this visual technique of accessing pictorial information will be analysed and discussed.

CHAPTER 4

SPATIAL INFORMATION SYSTEMS

Section no

THE PRIORITY OF SPACE AS AN ORGANISING PRINCIPLE IS SO COMPELLING THAT WE FREQUENTLY TAKE INFORMATION THAT MAY NOT BE CONSIDERED SPATIAL IN CHARACTER AND GIVE IT A SPATIAL REPRESENTATION JUST SO WE CAN THINK ABOUT IT MORE CLEARLY AND REMEMBER IT MORE ACCURATELY (Miller 1968).

4.1 INTRODUCTION

Within this chapter spatial techniques of analysing both textual and pictorial information will be examined. People use spatial techniques in their everyday lives (Fields and Negroponte 1977). In mentally travelling through three-dimensional space, spatial techniques are one of the tools used for finding one's way. Within a room a person is able to locate many objects not because of their names, but because of memory of the spatial locations of objects in relation to the rest of the room. Fields and Negroponte have suggested that provided the user has a degree of 'spatial familiarity' with the three-dimensional area in question then spatial memory can be applied to much larger collections of information existing in both real and virtual three-dimensional space. They have further suggested that such a technique may be applied to the accessing of data from both textual and pictorial databases.

Spatial navigation in such databases can be dynamic, that is, generated in response to the pattern of searches, or alternatively navigation can be based upon the known preferences of the database searcher. When navigating through virtual three-dimensional space, spatial analysis of information can provide a view of relationships that would be impossible in real space, and can even provide alternate spatial orientations to give the user literally a new way of looking at things! Spatial references within three-dimensional virtual space can be invented for just about any collection of information whether textual or pictorial, with options in some systems for the user to reorder the invented environment according to their taste or preference. This therefore provides an interactive technique by which data may be accessed and stored.

To be successful the interface must offer efficient methods of return when the user decides that one strategy is exhausted and they wish to explore an alternative direction. Furthermore the interface should make the underlying structure of the search process visually explicit so that the user can easily control the steps necessary to achieve their goals. Thus such an approach to accessing databases uses a combination of visual and spatial clues. This enables the user to navigate around the database without complex textual query languages.

Within these spatial systems users may therefore move around within the simulated space available in order to locate the information that they require. The most common strategy for such spatial information systems is the use of separate data management facilities for the spatial and aspatial description of entities, linked by common identifiers for entities. This reflects the concern that specialised data organisations and physical data structures are needed to achieve high performance in managing and retrieving spatial data, it is also the result of the historical separate evolution of aspatial and spatial data management systems (Kemper and Wallrath 1987).

Fields and Negroponte (1977) having suggested that data could be found spatially, were the first to develop the idea of spatial database systems in which clues about the data location and appearance could be used as well as its attributes. Within their theoretical system, data was stored in three-dimensional space viewed through a window. A user of their system would be allowed to operate a joystick in order to navigate through this three-dimensional information space in order to find the information that they require. Fields and Negroponte suggested that initially the user should be able to see representations of all the information stored within the database, however on approaching each piece of information, it should become clearer and more detailed. The user should also ideally have an appreciation of the spatial organisation of the information within the database in order to navigate effectively around it.

Fields and Negroponte further suggested that in order to help the user navigate around a very large database, navigational aids might also be required. Such aids might include both a different appearance in the form of the data depending on the user's spatial location within the system, and also general maps of the databases spatial layout, easily accessible at any spatial location within the system. Markers could also be used as 'signposts' to particular kinds of data, these markers could take the form of an object, or coloured or distinctive backgrounds within certain spatial locations. Additional navigational aids might include sound and movement from the data objects, the sounds becoming louder and more distinctive as the object is approached.

Fields and Negroponte concluded that spatial location of information is best for those problems in which the formulation of a precise query is difficult. The spatial technique encourages browsing through the data, allowing information to be found even when only a vague specification is possible. When a key can be precisely specified, it may be more efficient to type that key directly into the system rather than attempt to navigate through the spatial database. The ideal combination within a three-dimensional spatial database would therefore allow a combination of keywords and spatial navigation in order to specify a query. Practically it is suggested that spatial database systems can be broadly subdivided into two kinds. Those databases where the information stored within the database is of a textual nature but is accessed using a spatial technique, and secondly those databases where although the method of access is again spatial the information stored within the database is pictorial.

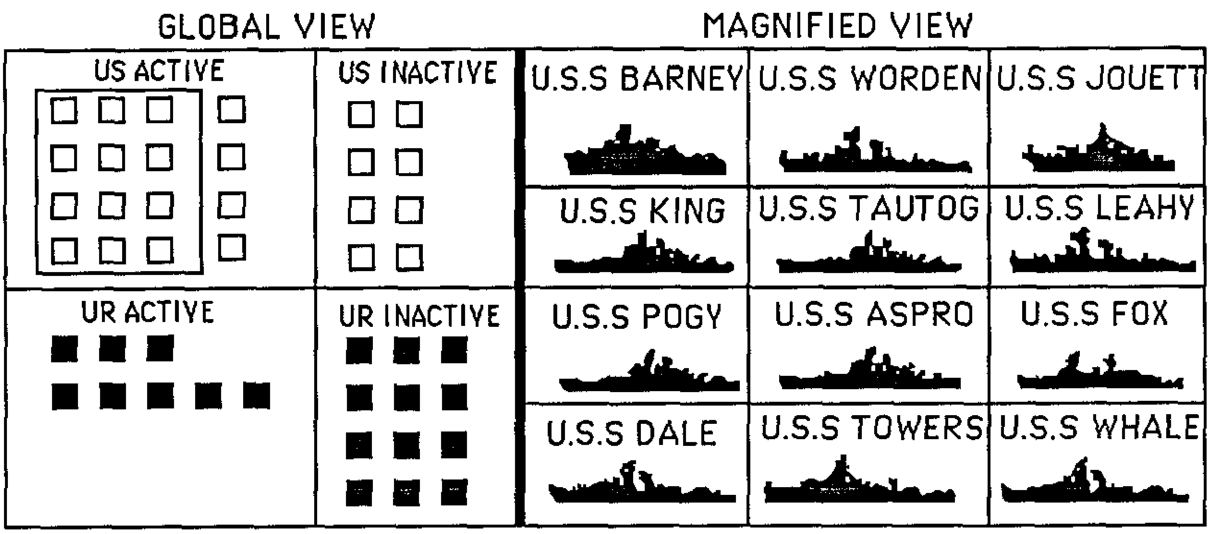
4.2 SPATIAL DATABASES WHERE THE INFORMATION STORED IS TEXTUAL

4.2.1 Herot's Spatial Database of Ships

One of the first database examples in which the technique of spatial navigation within virtual three-dimensional space has been applied was Herot's database of ships (Herot 1980). In this system information was displayed to the user within separate areas of a screen, the user then highlighted the area of interest satisfying the informational need of that moment. This information then directed the user to yet another location within a different screen. Here the virtual three-dimensional spatial concept was utilised by arranging the data within the system in levels of detail which progressed from the more general to the more specific. Thus the user viewed data at one level, chose from among available alternatives and on the basis of their choice moved from level to level to obtain the degree of detail they required.

Herot's system enabled the user to visually navigate around a database of ships, identifying the particular ship that they were interested then obtaining more detail by using a joystick to zoom in on it. Practically the system superimposed a graphical data space on an INGRES database allowing the user to examine the entire database by manipulating the symbols or icons in the graphical data space (Korfhage and Korfhage 1986). The two essential parts of the system were the 'global-view' map and the magnified-view of the data surface. The global-view map displayed all the ships within the database. The magnified-view displayed a section of the global-view map in greater detail. On using the magnified-view its location was displayed on the global-view map by a highlighted rectangle which could be moved around the screen with the aid of a joystick. Using this technique users were able to move around the display in any direction selecting which ever ship they required. Both the global-view map and magnified-view map are shown below.

FIGURE 4.1



THE GLOBAL-VIEW MAP AND THE MAGNIFIED-VIEW MAP

The rectangle around twelve of the US active ships in the left hand global-view represents the view shown by the magnified portion. The global-view was further separated into the two countries United States of America (US) and Russia (UR).

Rows within the global-view were subdivided into both active and inactive ships with individual columns indicating different classes of ships. Each individual ship was represented by an icon (a graphical representation of the ship in question) the size of each icon being directly proportional to the size of the ship it represented.

4.2.1.1 Using the System

In order for the user to highlight a particular ship the joystick was operated until the required ship was at the centre of the rectangle in the global-view map, the handle of the joystick was then twisted clockwise. This selected the view shown on the global-view map and magnified it so that further textual information such as the ship's name, could be seen. At this stage the user had three alternatives, they could move across the data surface and examine other ship's in similar detail, they could return to less detailed information by twisting the joystick anticlockwise, or they could twist the joystick clockwise again providing details about the ship selected such as the ship's speed and range. In addition to selection by icon, conventional textual query languages could be used to access information stored within the INGRES relational database enabling selection of specific data items from the database.

4.2.1.2 Functions of the Database

Herot's spatial database management system had several important features. The joystick used provides a simple and natural means of moving through the database. The zoom capability both provided an enlarged view of a database segment as well as allowing graphic portions of the icons to be defined at several different levels of detail so that zooming provided a more detailed and descriptive graphic. These graphical representations could thus convey information by colour, arrangement and highlighting. Similarly the textual information displayed could be keyed to the zoom, so that focusing in caused the display of data values that were invisible at more global levels of viewing. By concurrently presenting a global-view and a close-up view this system showed the user both the kinds of information available and selected sections of information at the same time. However the current version of the system unfortunately involves too much direct intervention of a professional (the database administrator) to set up the graphical data space. This means that although the system is easy to use and browse through, it is difficult to append new information to the database such as the introduction of new ships. The user is also unable to add additional attributes to those already present within the system.

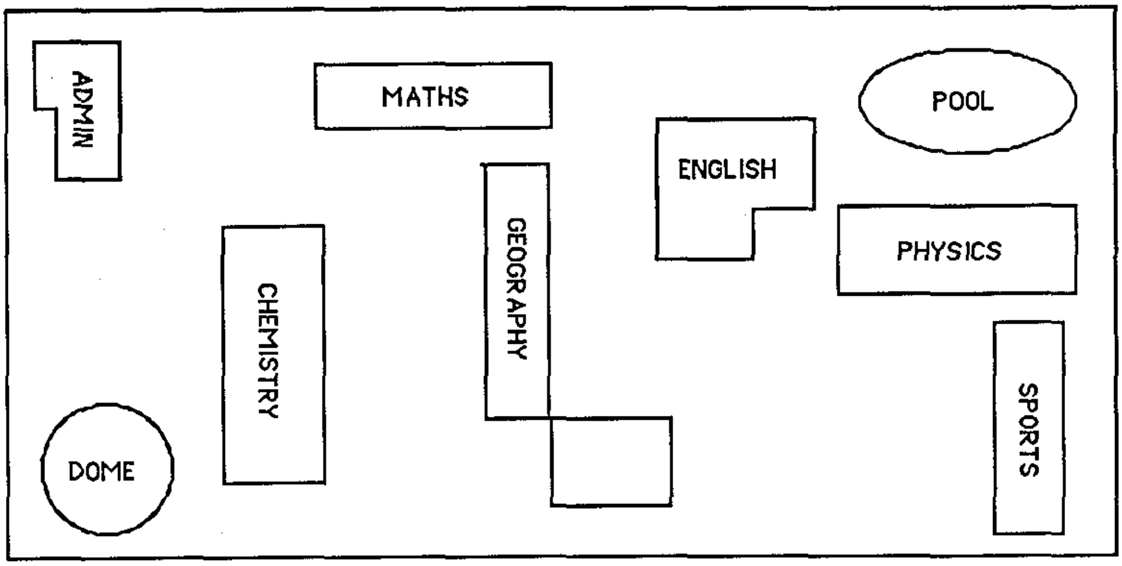
Herot's system shows that the principal of spatial database management can be applied to systems such as a database of ships whose textual information does not have an inherently spatial layout. Provided that the objects in question can be represented by photographs or textual descriptions and accessed by a form of iconic representation, then the principles of spatial database management and navigation within virtual three-dimensional space can be applied. By using different icons a wide variety of pictorial databases could be examined using a similar spatial technique of access. However within Herot's ship application the method of access employed is spatial in nature rather than the actual data itself. The most important advantage of the system is that that the user does not need to know the database organisation or query language syntax in order to use the database, relying instead on the ability to navigate through the conceptual three-dimensional space provided.

4.2.2 Filevision

One of the most popular software packages used by microcomputers in the analysis of spatial information has been Filevision (Nordgren 1986). Filevision is a software product available for the Apple Macintosh which has an ability to manage and access data which has both a high visual and spatial component. The main function of Filevision is the storage and retrieval of object information contained within a picture (more usually called a map in Filevision nomenclature). Using the Macintosh mouse, information can be obtained from such maps by clicking on objects of interest, users moving from one map to another by the selection of an appropriate icon. As with Herot's work, Filevision maps when linked correctly can allow the user to spatially navigate through a database in order to find the information that they require.

A simple Filevision map is shown in figure 4.2 for a university campus. Here different buildings are shown on the initial main map. These buildings act as icons which can be selected producing yet more detailed maps of the various floors within the buildings. A yet more detailed map of the rooms within those floors can then be chosen. Thus such maps are linked hierarchically in increasing levels of detail.

FIGURE 4.2



BUILDINGS WITHIN A CAMPUS MAP WHICH ACT AS ICONS

Icons arranged within each of these maps contain a type. This is the arrangement of data fields and information associated with the icon. Various textual search techniques can be applied to these data fields enabling the correct icon to be selected. These search techniques include the ability to select icon's based upon parts of their names, for example one such textual query based upon part of an icons name might be "select all buildings which end with the letter y" giving the answer chemistry and geography from the map above.

Filevision's strength is its ability to combine pictures with text on the same screen and the ability to select information either visually or textually. Icons visible on the screen can be selected by mouse clicks extracting more information about the object that the icon represents and textual queries can also be posed about the descriptions of the icons. Another significant advantage of Filevision is that it allows a large number of records in the form of icon's to be displayed at any one time. This enables spatial relationships to be distinguished simply by observing their appearance on the screen. This visual approach enhances the usability of such a system by enlarging the selection of information that is readily provided to the user.

Unfortunately the form of the textual query that the user can ask is restricted to those already contained within the system, there is no way that users can design their own queries by programming. Therefore the user is constrained to developing applications which suit these textual queries. Another problem is that a particular spatial arrangement of icons can not be searched for within the database. This prevents queries such as "find all buildings less than fifty metres apart" or "what is the building nearest to the chemistry building" being asked. A database system which allowed the user to find the spatial locations and arrangements of such icons would be a valuable addition to the Filevision system.

4.3 SPATIAL DATABASES WHERE THE INFORMATION STORED IS PICTORIAL

4.3.1 The U.S.S Carl Vinson

The principles and theories behind Herot's database of ships have been applied on the U.S.S Carl Vinson an American nuclear-powered aircraft carrier (Kramlich 1984). The spatial database system developed supported a facility for embedding computer programs as data types in its data surfaces. This information was obtained via 'ports' which enabled different subsystems to be accessed. Within this application there were two basic sub-systems, a platform display subsystem and a map display sub-system. Information from sightings, radar, sonar and scout aircraft were all constantly added to the system acting as a central repository for the incoming information. This information was in the form of sightings, photographic data, graphical data and static and numeric information and from all this data a picture of the current positions of both friendly and enemy craft was determined.

Within the map sub-system ship icons were displayed overlaid on a map of the world in order to denote their position. The maps were stored as photographs on a videodisk and the positions of the ship icons were updated as new information arrived. A user navigated on the maps just as on a conventional data surface, by means of a joystick. Thus the user scrolled around the maps horizontally and vertically and zoomed in to see a more detailed view. In addition to the basic zooming and scrolling functions, the map display sub-system provided the user with a menu of display options and a graphical editor for annotating the maps.

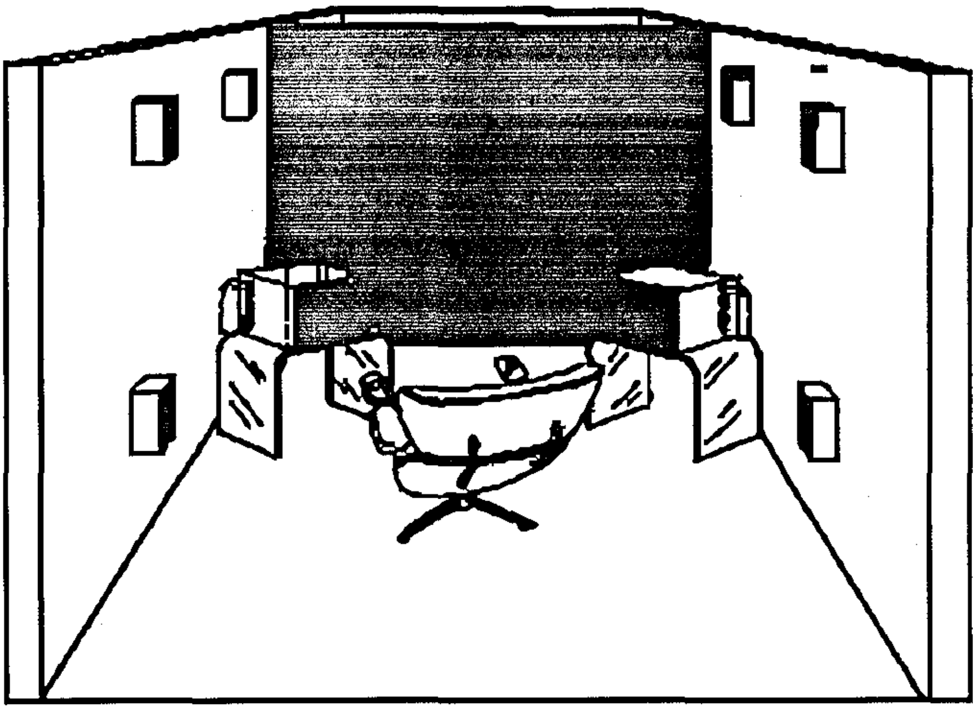
However since the information was stored on a videodisk the only way that new information can be added to the system was to purchase a new videodisk, which was impractical since there was a need to constantly add new information. A videodisk from which information could be added to and erased would be a considerable advance. In general, applications which entail the constant updating of pictorial information should where possible have storage media that allow such changes. The application above is highly interactive requiring constant analysis of information and thus the effectiveness of videodisks as a storage medium is therefore questionable. Videodisks should primarily be used where the number and nature of the images stored within such a system is static.

The U.S.S Carl Vinson system demonstrates many of the advantages of spatial databases. Users need have little or no knowledge about the physical storage of the information, instead relying upon spatial techniques to locate items of interest within the database. These systems allow users to access computer resident information visually rather than entering textual queries. Browsing is also encouraged enabling users to find information without having to specify it precisely. Spatial database management systems can in general be used to display many forms of data such as video images, text documents or computer program output enabling users to locate such information in a database by navigation through a three-dimensional virtual space.

4.3.2 The Media Room

Another important spatial database management system is that of the Media Room at the Massachusetts Institute of Technology called Dataland (Donelson 1978). This is a site in which the room itself is the terminal. One wall is completely covered with a screen so that a colour television can back project on it.

FIGURE 4.3



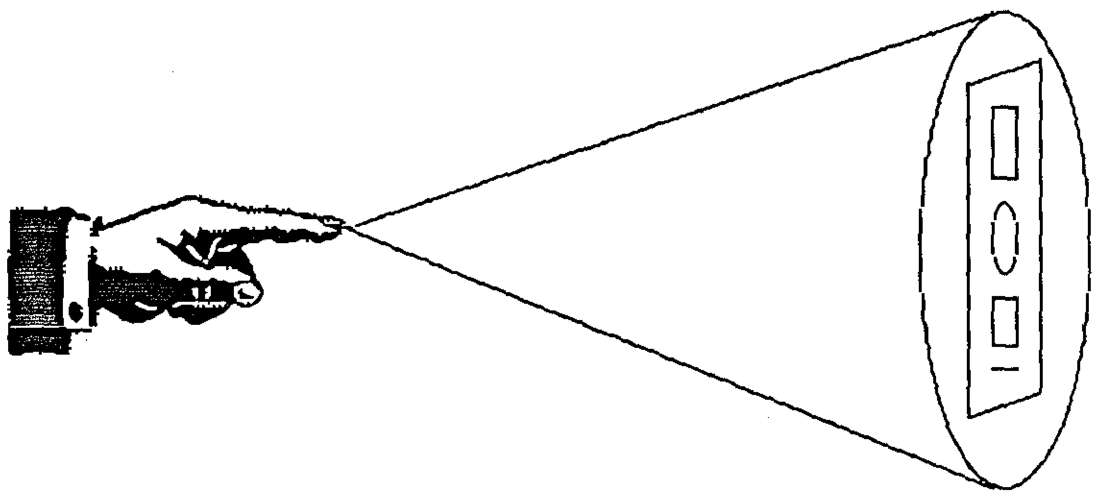
THE MEDIA ROOM END VIEW SHOWING FULL DISPLAY

The user sits in a chair in the room so that they can see all of the screen. The chair has on it two joysticks, one on each arm of the chair for easy access. Next to each of these joysticks is a square shaped touch-sensitive pad. On either side of the chair and within arm's reach are two television monitors with a touch screen facility on each. At each of the four corners of the screen in front of the chair is a loudspeaker as well as another four speakers behind the chair ensuring the user is surrounded completely by three-dimensional sound.

The user navigates around the Dataland surface with the aid of a joystick. Moving the joystick enables the small rectangular 'you are here' overlay marker on the world-view monitor to move in the same direction. Another method of travel within this system is via voice. Here the user tells the system that they wish to travel to a certain location on the spatial map by giving directions relative to objects, such as "to the left of", "above", "below" etc.

Yet another method of navigating around the Dataland system is by 'touch travel' on the touch sensitive surface of the world-view monitor. By touching the location that the user wishes to travel to, the large screen in the front of the user displays a view of the Dataland surface at the destination point selected. Users can also select items from the screen by pointing using the technique of 'cone of indication' shown below. Here the further away from the object of interest, the wider is the base of the cone. Therefore moving the pointing finger closer to the screen produces a smaller cone of indication sharpening the reference to a chosen object.

FIGURE 4.4



CONE OF INDICATION

The nature of an object within Dataland determines the form it will take. Zooming in on a picture of a book means that any chapter can be automatically arrived at simply by pressing the joystick at the position corresponding to a chapter heading that appears on the screen. Page flipping animation occurs on the screen as the user moves from section to section. Individual pages can be turned just as easily with the aid of the monitors. Similarly maps may be used within Dataland enabling the user to conceptually travel around the map's image.

4.3.2.1 Conclusions about Dataland

The researchers working on Dataland concluded that a spatial view was extremely advantageous but that it should not be much more than a 'sufficient' idea of space, a minimum indication that would cause each user to form an individualised internal sense of position and relationships, rather than a fully detailed space. They also concluded that items stored in Dataland should be easily recognisable. A book should look like a book and a telephone like a telephone. Users should not have to waste time learning what the visual symbols might mean or how to manipulate them.

One of the major problems with both this system and spatial database management systems in general, is determining the organisation of objects on the screen. These tend to become disorganised easily unless careful attention is paid to each item in turn. Allowing users to manipulate information around the screen often produces confusing spatial layouts of information unless great care is taken. Here there is a conflict between allowing users to arrange data to suit their own preferences, and producing an uncoordinated spatial arrangement of objects within the system which makes identifying their location difficult.

Since within most spatial databases there are a multitude of images present, some form of structure enabling users to easily access these images is required. However navigation through such structures is often problematic. As users navigates down a tree structure to more clearly identify an object of interest, they tend to become disoriented. In wishing to move from one photograph in a tree hierarchy to another photograph in a different part of the tree hierarchy there is no easy method except to move back up the tree structure and down the appropriate branch. Attempts at trying to rectify this problem within Dataland suggested mapping the tree structure of nodes onto a two-dimensional display screen. Initially some quite complex neural net-like schemes were developed but these were found to be difficult for users to peruse through. Yet where a simple hierarchical set of planes was used, there was difficulty in representing the often complex nature of the data, there were also many cases where the data did not lend itself logically in a hierarchical manner.

This problem to some extend remains unresolved in that there are two opposing factors to take into account these being the simplicity required in order to enable users to easily peruse data contained within a system and the complexity required in order to adequately represent the data in a logical fashion that allows users to understand how an item of data relates to the data surrounding it. This implies that careful consideration needs to be taken in the logical linking of items within a spatial database. The Dataland experiment reinforced the belief that people like to stick to familiar representations and quickly grasped concepts, enabling them to accomplish things as quickly and easily as possible

The most important advantage of such spatial database systems is that the user's ability to navigate within virtual three-dimensional space can be used as a means by which to access information. The information accessed can consist of anything that the user wishes it to contain such as images of maps, personal data, letters address books etc. The system however allows the user to peruse the data surface both quickly and efficiently locating data within the database particular to the user's own personal spatial layout.

4.3.3 Yosimura, Tanaka and Icikawa's Visual Interface for Geographic Information Retrieval (1983)

This pictorial database consists of sections of detailed information about parts of a geographical map, such as the road network and the features and public utilities present. These can be superimposed on each other enabling the user of the system to see these combinations of data to their best effect. The geographic information retrieval system within the database allows for the semantic zooming and panning of displayed images. Zooming allows part of the database to be increased in resolution. Panning allows the user to move across the database in order to obtain the correct image, after which the image can be zoomed if required.

Overall the four facilities that are provided to enable the user to work with this information retrieval system are:-

1) Graphic Zooming

This allows the user to scale in or out in an area of current display.

2) Graphic Panning

This allows the user to shift the coverage of the area to be displayed.

3) Semantic Zooming

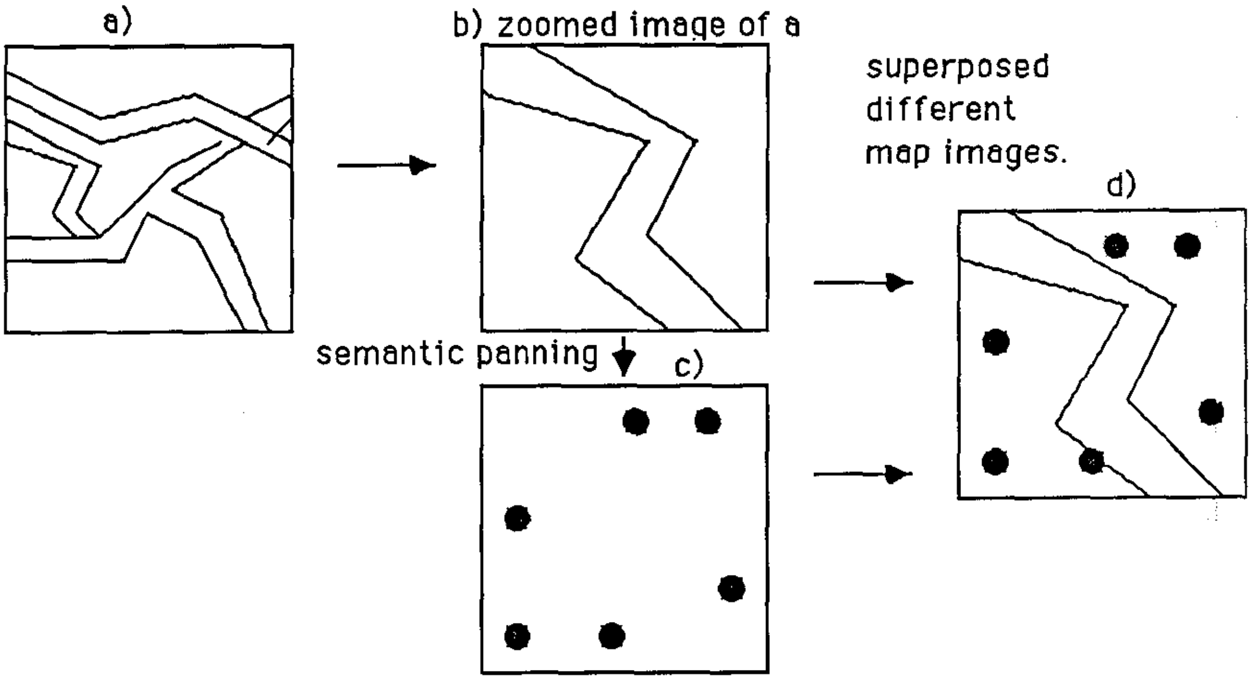
When categories in a logical hierarchy are displayed, zooming in on them changes the function. For example semantically zooming a forest might produce a category of tree types within that forest.

4) Semantic Panning

This enables the user to select from a logical hierarchy of categories by moving across the display to select new information. For example in the diagram for this section the user could semantically pan from roads to railways.

For example:-

FIGURE 4.5



EXAMPLE OF SEMANTIC ZOOMING AND PANNING

Here a) is the image of all the roads in the area

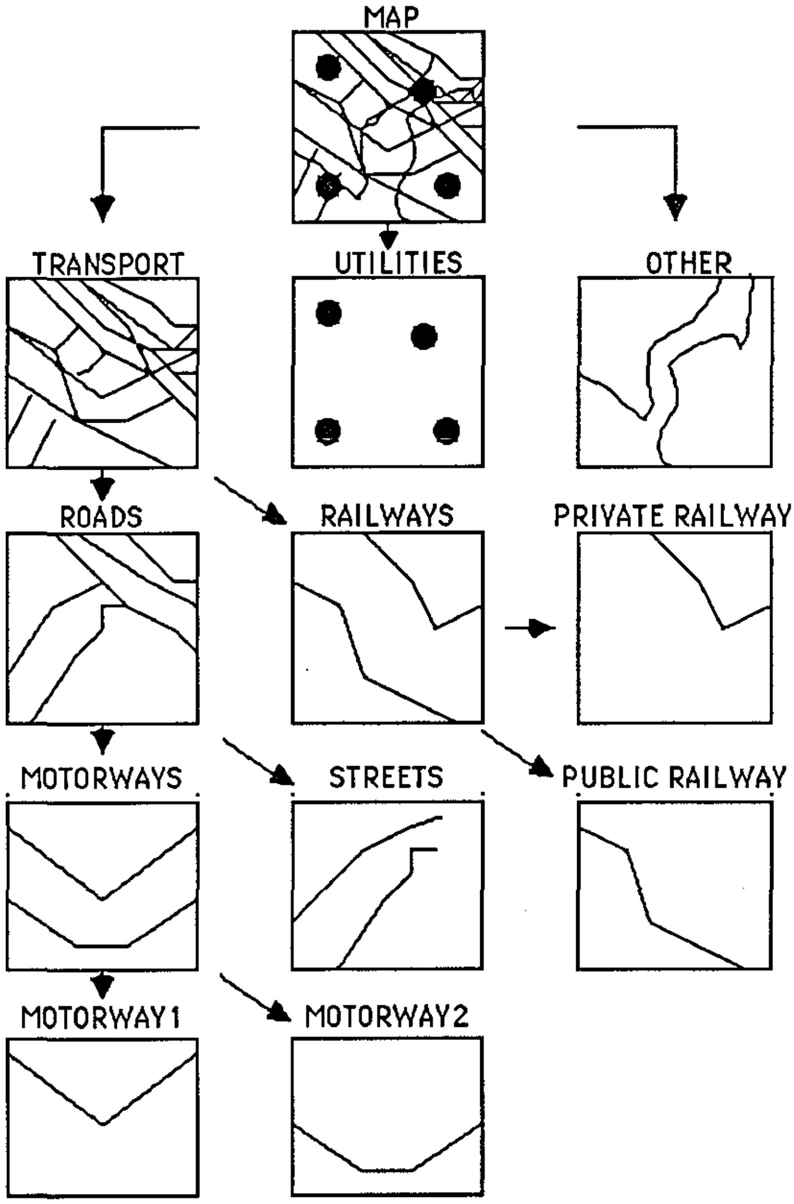
b) is a zoomed section of the roads in image a)

c) is the location of important areas within image b)

d) is the superimposed images of b) and c) which enable the user to identify important areas on the roads.

Thus for a complete map possible zooming and panning might be

FIGURE 4.6



EXAMPLES OF LOGICAL HIERARCHY ALLOWING SEMANTIC ZOOMING (VERTICALLY) AND PANNING (HORIZONTALLY)

The major advantage of this system is its ease of use. This is because a simple pointing device can be used in order to both zoom and pan around the pictorial database. The system operates by allowing the users to search the database using their spatial knowledge of the area. Therefore provided that the user has some familiarity with the area they are searching, the database is able to assist in providing more detailed information.

Such systems are however at present exclusively used for geographical maps. Initially a database designer often has to identify semantically similar objects within images to allow semantic panning, since image analysis systems have great difficulty in distinguishing between objects within the images such as motorways and roads. Information must therefore be manually extracted from the raw image before it can be used within the logical hierarchy system. Problems with the system include the fact that since it consists of a tree structure, the addition, subtraction and replacement of images is not easy. Such alterations may be required where new roads, railways or other such features have been built and this requires adjusting all the logical and physical maps within the altered objects hierarchy in order to take account of such changes. Also unlike many previous pictorial database systems no query facility exists whereby the user can textually ask the system for information about the database. Thus textual queries such as "find me the M4 motorway" are not possible within this system because it has no method by which to identify such an object. Users are therefore limited to visually finding their own way around the geographical area.

This system demonstrates that both zooming and panning can be used in a pictorial database. It relies upon the ability of the user to spatially navigate around a geographical area and is most effective where the user has a degree of spatial familiarity with the area in question. Since within this system users have been found to have no difficulty in panning across related geographical pictorial information, this ability to pan across pictorial data may therefore be useful in specifying visual queries as well as selecting information.

4.3.4 Spatial Movie Maps

Another spatial database application is Lippman's Movie map (Lippman 1980). Aspen, a small town in the United States of America has been filmed with videocameras and the film stored on a videodisk. An interactive map has been designed and built to enable the user to undertake a simulated drive around the town by manipulating the information on the videodisk. The user of the system may control the speed, route, angle of view and mode of presentation of the information and thus receives a simulated view of a tour of the area. At its simplest level this system may be regarded as a replacement for a road map, although the increased interaction of the user via user-controlled turns, speeds and points of view ensures a much greater familiarisation with the area than would be possible using a simple map.

A second purpose of this system is to provide a data access and information system about objects and buildings present within the town. The nature of the building determines the data contained within it, therefore within a restaurant lies the restaurant menu for the day, within the town hall lies all the towns vital statistics and historical records, and the telephone exchange contains the town's telephone directory.

4.3.4.1 The Basic Map

The videodisk used to film the town had the ability to store and randomly access 54,000 individual video frames. The videodisk map gives the impression of motion by displaying a set number of frames as the user conceptually moves through the streets. During the user movement around the town a pointing device was used for driving control. Driving control possibilities were displayed on the screen where the apparent movement took place, this allowed the user to change the action of the videodisk, such as stopping it or running the car in the reverse direction. All these motion possibilities were also available with a joystick, the joystick also being used as an accelerator when moving around the town. Sound was also available generated from a speech synthesizer, these sounds consisted of compass directions, landmarks and names of streets when approached by the user.

However as with all videodisk systems Lippman's application suffers from the inability to change the nature of the information stored. New buildings or the demolition of old buildings means that a completely new videodisk needs to be pressed in order to incorporate the new information. Movie maps in general encourage spatial methods to be used in the analysis of information. These systems attempt to use a subject's spatial ability to navigating around a town rather than specifying locations based upon their names. They simulate real world movement in a realistic and effective manner in a way that the user would encounter if they were actually at the town displayed on the videodisk. Movie maps have both a spatial method of access to data and data that is inherently spatial in its nature. The spatial method of access is realised by the conceptual navigation through the videodisk, while the spatial data is realised by being a geographical example. The most useful purpose for movie-maps is in providing users of these systems with a knowledge of a location they are unfamiliar with, in addition to the ability to conceptually navigate both easy and quickly around the area. Movie maps suggest that systems which allow users to navigate and manipulate information within a conceptual spatial system should where possible be as close to the user's idea of reality as possible. Thus such systems should be constructed in a form which allows users to navigate and manipulate information as they would in the real world.

4.3.5 <u>Interactive Videodisks, Real and Imaginary Spatial Landscapes</u>

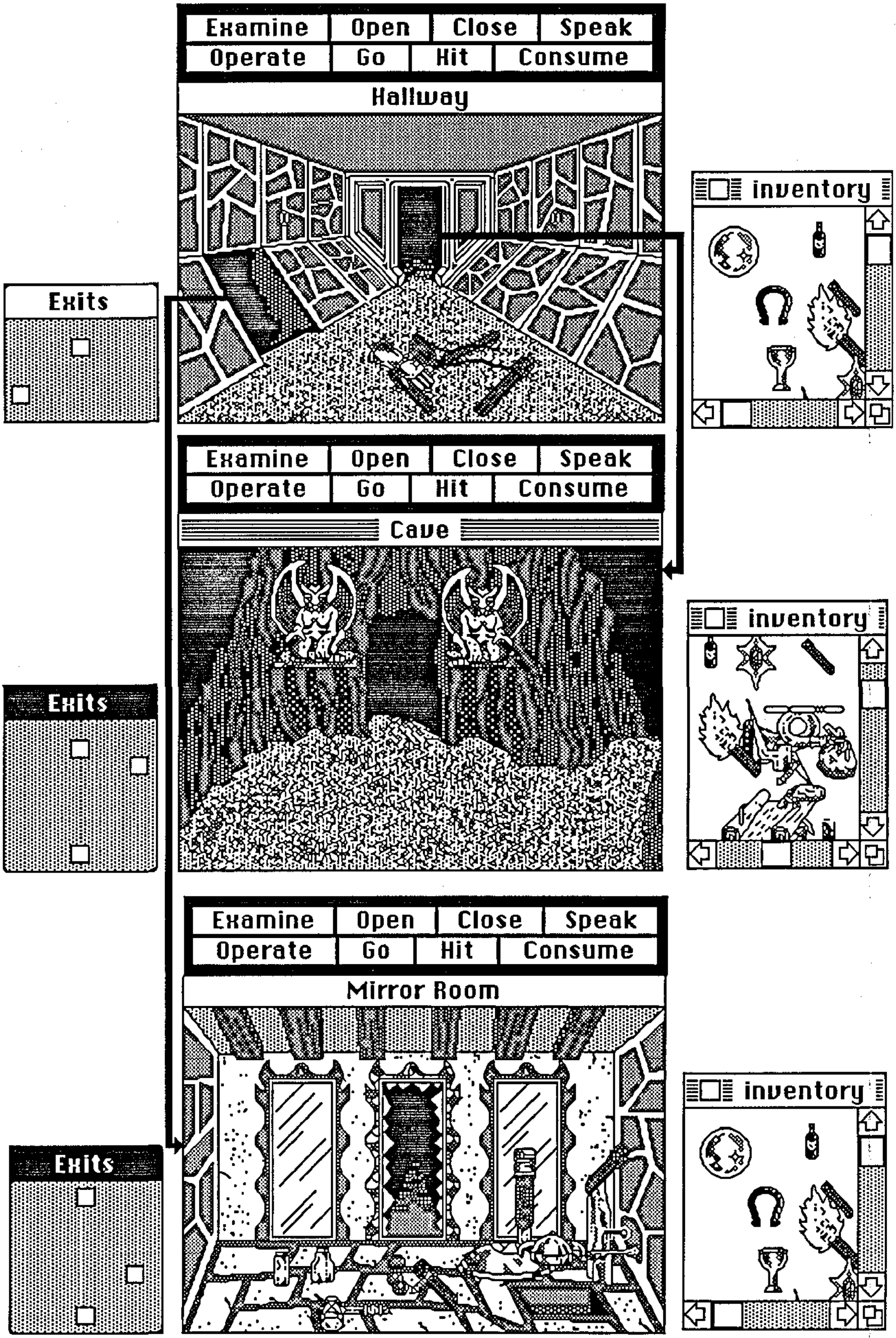
Moving beyond the perception of a real geographical location, interactive videodisks have been created that use real physical locations but which are assembled in a fictitious way (Veith 1988).

The Ford Motor Company have developed an interactive videodisk system which was intended to help car dealers become familiar with videodisk technology. Here the disk was essentially a golf game. The videodisk displayed a person, showing only the shoes, walking up to a tee. Using graphics, the player was asked to select a club and determine the amount of force to use, how to swing the club, and other details of a golf stroke. The video then showed the ball being hit and depending on the details chosen, the simulated golf ball landed on the fairway, in the rough, or in another virtual location. Each of the 18 holes within the course were taken from real life golf courses and were chosen because of their memorability as representations of a hole for those courses.

The geographical location can also be taken beyond the mixing of real places and can be entirely fictitious. The most common example of such systems are games based on interactive videodisks. A game with spatial dimensions creates a logically consistent geography for the players, even when the geography includes black holes, time warps, trapdoors and other procedures for being in a different location. Figure 4.7 is taken from the Apple Macintosh game Shadowgate and shows this principle. Here three areas are spatially linked by exits that the user can move through.

Although both the golf game and Shadowgate modify reality, users are still able to understand and use the spatial organisation of the information provided in order to locate information and navigate around the systems. This shows that a system may not necessarily need to be within the user's experience for them to navigate successfully around the conceptual space.

FIGURE 4.7



SHADOWGATE GAME IN WHICH AREAS ARE SPATIALLY LINKED

4.3.6 Conclusions about spatial database systems

Spatial information systems provide novel opportunities for decision-aiding. In these systems humans manipulate information both spatially and linguistically, these two modes are complementary and together provide a power that neither alone can achieve. Dialogues can be described in terms of a layered protocol whether they be between humans, between computers or between a human and a computer. An intelligent interactive spatial information system should incorporate some intelligence both in its control and dialogue and in its interpretation of the data that it contains, thereby providing support to its users and where possible warnings about inconsistent data.

Many of the spatial systems discussed in this chapter incorporate selection of visual or textual data by using mice and the ability to pan and zoom a picture enabling users to more easily browse through the database than by textual techniques alone. The ability to 'zoom' selected representations of data suggests that where data can be decomposed into its constituent objects, then a hierarchical structure may be the most appropriate. The database system can again assist in providing more detailed information when panning across pictorial information.

Graphical capabilities of these systems convey pictorial information by providing detailed graphic representations (icons) which allow users to easily select the information that they require. Such icons should where possible be easily recognisable. This selection of information by icon means that users do not need to understand the structure of the pictorial database in order to find the requested information, many such icons can also be specified together often allowing the user to visually choose an appropriate iconic representation from a screen menu. These iconic representations appear to be a useful technique by which pictorial information can be located.

Ideally interactive spatial databases should provide the ability for users (rather than programmers) to update and include information both easily and quickly. An interactive spatial database should at the very least provide the users of the system with a interactive environment which will allow them to query the system about the nature of the data stored within it. An expert system may be necessary here.

The interface should be as simple to use as possible rather than using access techniques specific to a particular domain. The method of access to this data depends largely on the nature of the data being analysed. Hierarchical menu systems may well be appropriate where the data can be decomposed into constituent objects thereby allowing the user to 'zoom' through ports to obtain the required information. Spatial database management systems are primarily concerned with techniques for locating pictures rather than detailed descriptions of the pictures themselves.

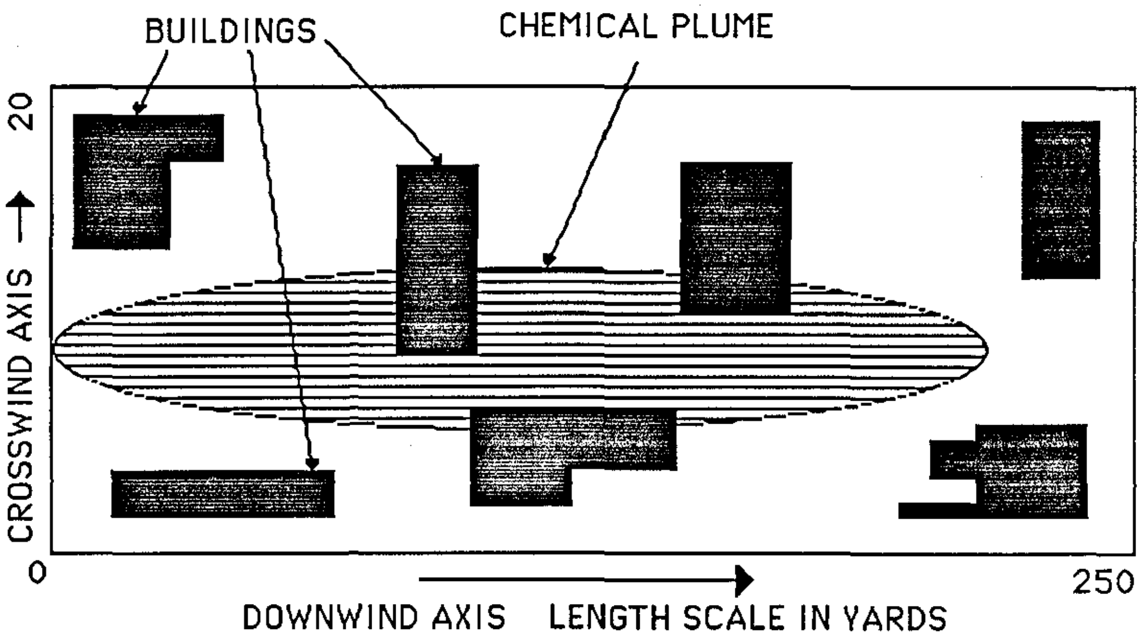
All the examples within this chapter so far have relied upon a subject's ability to spatially navigate in virtual three-dimensional space either in a real or imaginary domain. These systems have made no attempt to use spatial techniques to analyse information within pictures. The third category in this chapter containing two examples, the Gould, Tatham and Savitsky chemical emergency management database and Tranowski's scene analysis application are both systems in which spatial techniques have been applied directly to objects within pictures rather than as a technique to locate pictures from within a database.

4.4 SPATIAL ANALYSIS OF PICTURES

4.4.1 Spatial Techniques For Chemical Emergency Management

Spatial mapping systems have been used to assist chemical industries in identifying hazardous chemical storage locations (Gould, Tatham and Savitsky 1988). Furthermore by adding map data for areas surrounding a facility and by utilising air dispersion modelling software, individual facilities have been allowed to assess the impact of chemical releases upon the local community. A key function here is spatial searching which allows a user-defined graphic area to be queried with the enclosed objects linked by their geographic coordinates to an attribute database. In the event of an emergency due to a chemical spill it is helpful to know which areas are immediately endangered. Queries such as "all hospitals and schools in the path of a given plume for the release of a given hazardous chemical" can then be obtained by querying the system. These systems exist by allowing information about local streets, railroad networks and population locations to be overlaid upon the plume of released chemicals. Textual queries about the intersections of these two maps can then be executed in order to identify potential health problems. Queries can be asked such as "How many people under 5 and over 65 are within the intersection of the plume and population area, and which roads are available for evacuation". The graphical representation of such a system is shown below.

FIGURE 4.8



GRAPHICAL REPRESENTATION OF OVERLAID CHEMICAL PLUME AND POPULATION AREAS

Building a comprehensive chemical emergency management system consists of at least four separate yet integrated components:- a mapping package, an air dispersion modelling package, an industry related (chemical) database and a database management system tying these together, the database management system can be considered as the 'shell' under which the components operate.

This system extracts information by overlaying maps and thereby enabling textual queries to be asked about the intersectional information obtained. The only spatial relationship is that directly obtained by the intersection of the chemical plume and population map. Its main purpose is in allowing textual queries to be posed to a specific two-dimensional spatial area in order to obtain information about the likely dangers of the plume. Therefore the general applicability of such a system outside of this restricted domain is questionable, however it does show that intersections of object's spatial information within pictorial systems can be used to identify important information.

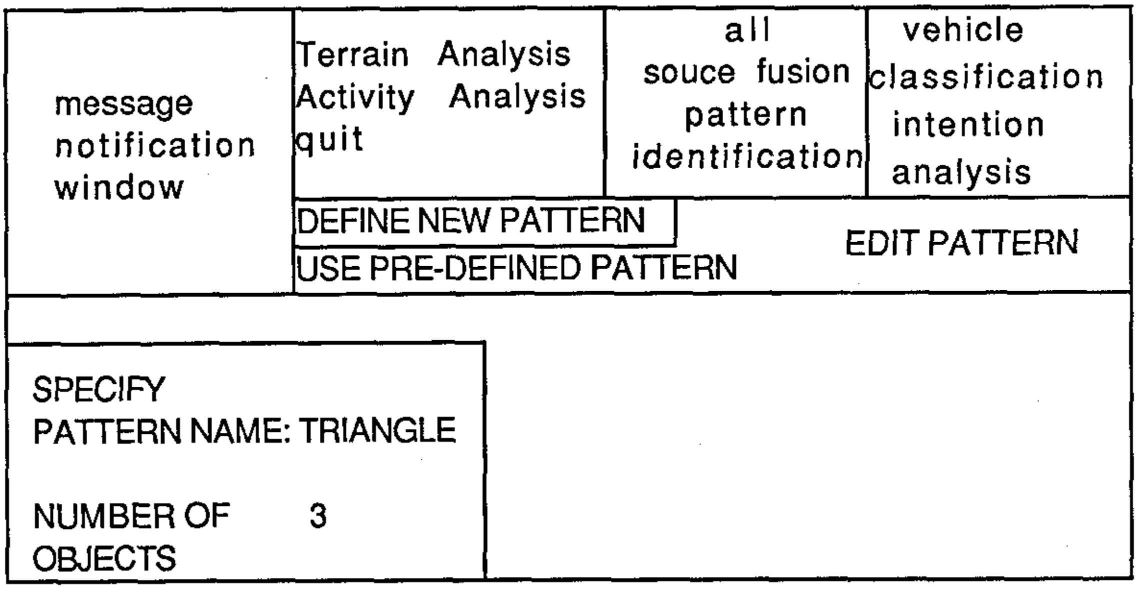
4.4.2 Tranowski's Knowledge Acquisition Environment for Scene Analysis

This system describes a knowledge acquisition environment used in the analysis of military scenes from aerial imagery (Tranowski 1988). It attempts to allow users of the system to analyse objects and manipulate the domain knowledge directly. "A perceptual and reasoning activity is of increasing importance in the capability to interpret and analyse imagery to detect objects and to notice changes in scenes" (Tranowski 1988). Spatial pattern identification is used in Tranowski's system to determine whether a set of objects derived from images are deployed in a pattern that is significant.

Within Tranowski's system users were allowed to enter both spatial patterns as well as describe properties of the patterns. Such properties include number of objects, the distance between each of the objects and the type of objects within the pattern. Tranowski's system allows the user to enter spatial patterns in two ways, either by selecting a pre-defined pattern or by adding new patterns to the system. Each of the objects can be moved anywhere within the working area and the pattern (along with the angles between the objects) is then saved. Once saved, any pre-defined pattern specifications with the same name are updated. A second mode of spatial pattern identification utilises one of the pre-defined patterns to describe the spatial layout of a specific unit. The interface provided by Tranowski for such an interaction is shown in figure 4.9.

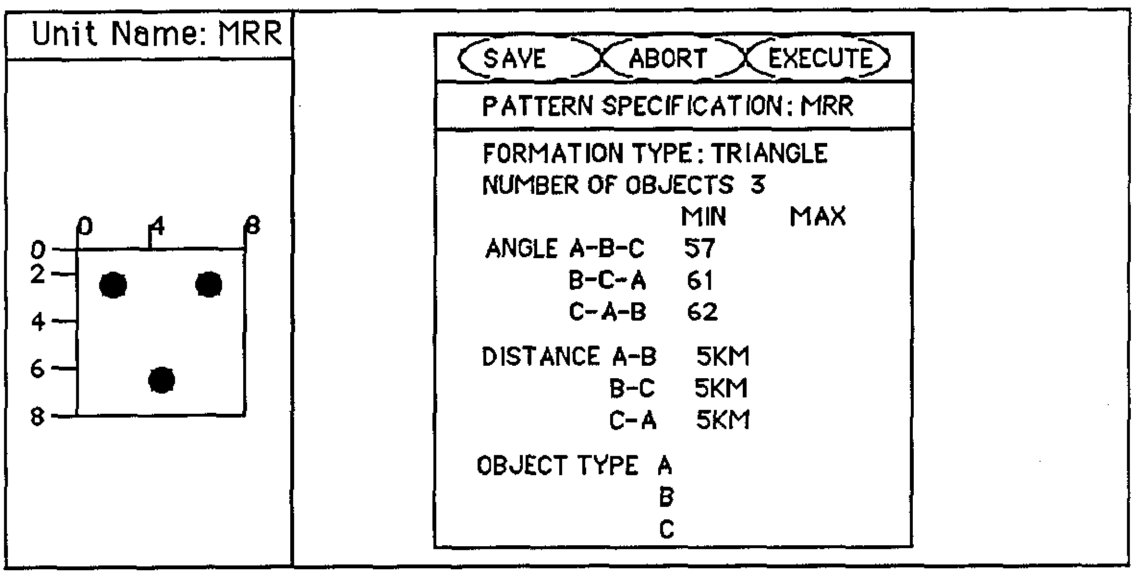
As an example in figure 4.10 a triangular pattern has been chosen, and the user is prompted to enter the name of a unit which has this triangular pattern (in this case Motorised Rifle Regiment (MRR)). In this way the MRR inherits the properties of the triangular pattern chosen. This pattern is shown in figure 4.10.

FIGURE 4.9



INTERFACE FOR OBJECT SPECIFICATION

FIGURE 4.10

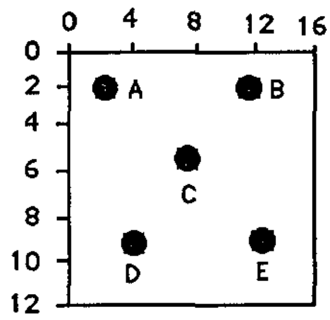


TRIANGULAR PATTERN SPECIFIED FOR OBJECTS

Here the text display has additional fields allowing the user to specify ranges for angle and distance as well as indicate the type of objects that compose the unit. The

angle and distance fields can also be edited. Modifications can be made by clicking the mouse on an entry in the text display, entering a new value upon which the graphic display changes. Once the user is comfortable with the unit's spatial description then the text and graphic descriptions can be saved for later recall.
In conclusion Tranowski's system stores groups of objects in a two-dimensional pattern which are available for later recall. However within this system one of the major problems is the situation in which more than three objects are present. Storing information and distance values then becomes difficult. An example containing five objects is shown in figure 4.11.

FIGURE 4.11



FIVE OBJECTS PRESENT WITHIN A PATTERN

Under these circumstances storing all angles and distances between five objects requires considerably more spatial relationships (consisting of angles and distances apart between the objects) to be stored. This makes the storage of all the distances and angles for situations with more than four objects impractical.

Secondly this system makes no attempt at allowing users to visually specify an object arrangement that they require, rather the system stores the object patterns and later recalls them based upon the name given to the pattern rather than allowing the user to specify the pattern visually by arranging objects on a screen. The user can only enter new spatial information by firstly recalling a pre-defined set of patterns and altering this information for the updated situation or by adding new patterns. There is no facility by which these patterns can be compared with the user defined query. For example a query such as "find me the closest pattern in the system to the one I have drawn on the screen" is not a query that

Tranowski's system can handle at the present time. Therefore this system relies upon the user's visual memory and perception in entering new spatial layouts of objects rather than locating specific object arrangements by comparing the user's spatial layout of icons with angles and distances apart of spatial information stored within the database. This restricts users to locating arrangements of objects stored within the system based upon their name rather than by using a visual query.

Both Tranowski's knowledge acquisition environment and Gould, Tatham and Savitsky's chemical management system rely upon spatial techniques for the storage of data within a picture. Tranowski's system however does not allow retrieval of information based upon spatial arrangements of objects within such pictures. Gould, Tatham and Savitsky's allows intersecting information between two pictures (the chemical plume and a map) to be identified but does not allow individual pictures to be selected based upon the spatial arrangements of objects contained within them.

4.5 CONCLUSIONS

Current textual query languages are designed as interfaces for a specific data model and cannot be easily accommodated to other models. SQL (Structured Query Language) for example, is the most common interface for the relational model and consists of tabular representation of relations, making the intrinsic form of the relational model visible to the user. While this approach is appropriate for applications in which users deal with tabular representations, it is suggested that in many instances such an approach is unsuitable for applications in which users do not have a mental model of information contained within tables such as in the case of pictorial databases. Spatial data is that whose relationships depend on their location rather than on their function. (Egenhofer and Frank 1988). It is suggested that visually object-oriented spatial query techniques enable users to perceive complex objects in a form that is closer to human's imagination and thinking than tables are.

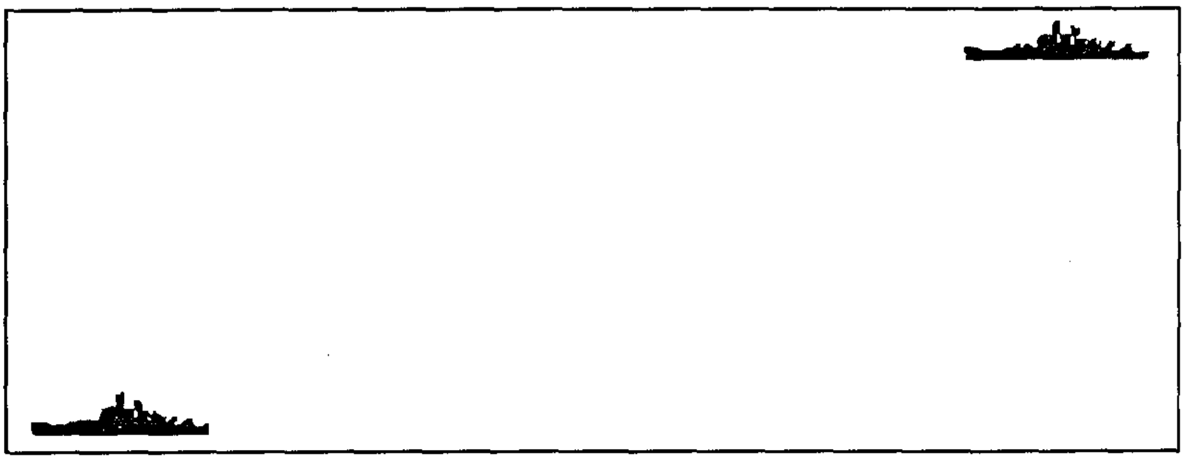
Spatial database management systems deal with large, heterogeneous collections of spatial and non-spatial data which require flexible methods for interactive inquires and representation. These systems offer significant advantages over conventional keyboard-oriented systems. Within these spatial systems graphics are normally used to convey information, this allows more freedom of presentation, information can be organised into trees, hierarchies or networks through which users can visually select their item of interest without a detailed knowledge of the structure of the database, thus there is a reduced requirement for subject training in order to use these systems. Simple concepts such as panning and zooming are also often provided to help the user to navigate through such systems. Zooming is particularly important since this allows users to become increasingly specific about the nature of the data that they require. Additionally panning allows users to move over picture surfaces to find information that they require. Visual browsing is encouraged within many of these systems since such browsing actions are easily reversible. This in turn enables novices to learn basic functionality within such systems. Such an activity is difficult within a conventional textual database management system where each piece of data must be requested explicitly.

Spatial database systems rely upon the user's spatial ability in navigating about a database. Objects (usually represented by icons) can be selected from such databases simply by pointing to them either physically, or via a pointing device such as a mouse. These iconic representations of data are (where possible) readily apparent in their data representation allowing such object representations to be selected visually rather than by using textual techniques of selection. Often such iconic representations enable users to explore the entire database with only simple controls, making the contents and structure of such systems easy to understand. Therefore locating information is simple and natural and the level of abstraction at which data is examined is similarly easy to control. Unlike conventional textual databases where users must pose successive queries, spatial database management systems allow users to explore the data surface in search of information. Finding such information is usually facilitated by the distinctive appearance of the data or the route used by which the data is accessed. Therefore it is suggested that such systems are fundamentally more cognitive than non-spatial databases where textual techniques are required to identify information. The actual nature of the data stored within these spatial database systems is not restricted to numbers and character strings. Spatial database management systems can present information that originates as textual documents, video images, computer program output or digitised pictorial information.

However these spatial database systems fail to extract data based upon how objects are spatially arranged within a picture, rather they use a spatial technique to access such pictures. Systems in this chapter do not attempt to locate pictures within a database by matching to the user's visual query, instead they either allow the user to spatially navigate through a database or use textual query methods to locate pictorial information. Thus a spatial visual query such as that specified below in which the user can spatially select and manipulate icons within an interface to represent the required spatial location of objects required within a picture can not be specified within present day spatial database systems.

In the spatial visual query below the user places iconic representations of objects on a screen after which the database management system is then asked to attempt to identify information stored within its pictorial database which matches to this user specified visual query. Thus in the query depicted the user requests all pictures from the pictorial database which have a ship both in the bottom right hand corner of the picture and in the top left hand corner of the picture.

FIGURE 4.12



USER'S SPATIAL QUERY ON AN INTERFACE FOR THE POSITIONS OF SHIPS WITHIN A PICTURE

It is suggested that this visual technique of picture selection may be applied to identifying pictures by specifying the objects within them rather than simply selection of pictures by hierarchically moving through pictorial menus.
Another problem with many of today's spatial database systems is the inability of the user to update the information contained within them. Because pictures are often linked in a hierarchical manner, subtraction or addition to the system usually requires a database administrator rather than the user to alter the information and provide the logical links between this pictorial information and existing data within the database. Thus users constantly need to rely on an administrator to modify the system. It is suggested that where possible pictorial database management systems should be developed which allow users to modify databases for both the incorporation of new data and the removal of old data. Thus although concepts such as iconic representation, movement through pictorial database by mouse control, manipulation of data in a life-like manner, panning and zooming and graphics to convey information all appear to be excellent techniques for locating pictorial information, these concepts have been primarily applied to navigation within three-dimensional space rather than as techniques for the locating of specific pictures within a pictorial database. It is suggested that these techniques along with other characteristics of spatial database management systems can be applied to two-dimensional picture identification as well. Such techniques should allow users to identify pictorial data by using computerised simulations of 'natural' methods which they would use in the real world to locate information.

It will be shown in the next chapter that both the spatial location and spatial arrangements of objects within a picture are memorable. It is therefore suggested that where spatiality is inherently involved in the data then the visual approach of object selection might be useful in identifying pictures. This identification may be based either upon where the objects are located within a picture or their positional interrelationships with other objects in a picture. It is further suggested that many pictorial databases might benefit by visually querying their spatial layout, even if this were but one of a combination of visual techniques which could be used to locate data. A system where pictorial databases are accessed by matching information stored to users visual queries on a screen could assist in enabling specific pictures to be located both easily and quickly, rather than forcing users to search through a pre-defined textual system. The psychological evidence for such a proposal is examined in the next chapter.

CHAPTER 5

THE APPLICATION OF PSYCHOLOGY TO PICTORIAL DATABASE SYSTEMS

5.10.1 Conclusions about Spatial Relationships within Pictures

5.10.2 Conclusions about Pictures

5.1 INTRODUCTION

In dealing with pictorial information, the important questions that need to be asked are:-

1) What is known about mental imaging?

Intuitively the ability to form a mental image objects and scenes will play an important part in identifying objects and constructing spatial searches for these objects.

2) What things are best shown pictorially?

Objects within a database may be better expressed pictorially than textually. Certain objects and shapes may be difficult to represent in the form of text, especially if they contain a complex arrangement requiring detailed textual specifications for a complete description.

3) What do we remember about pictures?

If certain visual attributes of pictures are always remembered, it would seem reasonable to use these attributes to locate pictures from a pictorial database.

5.1.1 What is a Picture?

A picture can be defined as a light reflecting or emitting surface manufactured by human or machine, and specified by a human with the intention of representing something (Scrivener and Schappo 1988). In general the real value of a picture is that it communicates facts about the things it represents. Complete understanding of a picture can only be accomplished if the subject has knowledge of the objects and relationships within such a picture.

5.2 IMAGERY AND PERCEPTION

5.2.1 Mental Imaging

Mental imagery refers collectively to mental images and to the mental processes which manipulate and inspect them (Julston and Baron 1985). Therefore imagery can be seen as the means by which a picture is stored in memory (ie brought into conscious attention). Mental imagery is important in the analysis of pictorial information because of the requirement to understand how images are stored within the brain thereby helping in the design of new object representations which are easier to understand in the subsequent locating of pictures.

Mental images are essentially pictorial and geometrically similar to the projections of the objects that they represent, they therefore have colour, brightness and depth. When a scene is projected into the brain, the resulting mental image is a representation of that scene. If the brain surface is composed of receptors which encode the brightness of each part of the projected image, the pattern of activity generated by the receptors is called a visual image (Neisser 1968). If the visual image includes the spatial locations of each object within it, the resulting representation is called a spatial image. Where depth is present within such an image it is said to be two and a half dimensional.

Subjects best remember images which refer to physical objects. This 'concreteness' is therefore the most important factor in the recall of an image, abstract concepts are much more difficult to visualise. This was verified by Paivio (1971) who asked subjects to learn sets of noun pairs, such as pencil/tree, soul/potato etc. Paivio concluded that subjects found it easier to form images of concrete nouns. Abstract concepts such as soul, health, pain and fear can be encoded only verbally, whereas nouns such as chair, table, car, house etc, can be encoded in both a verbal and an imaginal system and thus these 'concrete' nouns present greater opportunities for meaningful visual spatial associations.

However people's ability to visualise can be diminished if visually interfering tasks are presented at the same time. Atwood (1971) showed this by asking subjects to visualise a scene. Immediately following this, subjects either saw the numbers 1 or a 2. They were then asked to give the number that they did not see. For example if they saw the word "two" then their answer should be "one". Atwood demonstrated that visually presenting the 1 or 2 interfered with the subject's ability to visualise concrete objects. He also found that verbally presenting the words "one" or "two" interfered with visual tasks that involved abstract ideas. From results of this kind it can be concluded that there are at least two reasonably independent 'processing systems' within our brains, these are a visual system and a verbal system.

5.2.2 Perception

In order for its movement to be regulated by the environment, a living animal must be able to detect structures and events in its surroundings. This ability is called perception (Bruce 1985). People are usually able to perceive what pictures represent though they may never have seen the objects within the pictures themselves. Even when presented with pictorial information they do not recognise, people are still able to perceive some structure to the picture although they may not understand its meaning.

This ability to be able to 'understand' pictures is due largely to our learning through experience and instruction when young (Deregowski 1974). Therefore in situations in which alternatives are possible, visual perception selects the most likely conclusion under the current circumstances from among the available possibilities (Murch 1973). Having 'understood' that a picture is a flat representation of a three-dimensional object means that the perception of a picture is very similar to the perception of a three-dimensional object, therefore picture perception can be approached from the point of view of visual perception and techniques used in the analysis of three-dimensional objects can be used in the perception of pictures.

5.2.2.1 The perception of grouped objects

Perception usually involves the understanding of a number of complex patterns. Fortunately interaction with the environment does not take place in a serial manner. Wertheimer and others (1958) outlined the Gestalt concept of pattern organisation in which perception involves the organisation of sensory input into units rather than the processing of separate sensations. Their idea was that there is instantaneous organisation when subjects encounter objects for the first time and that the spatial and temporal relationships between such objects is as important as the objects themselves. The human brain therefore tries to produce patterns from random selection of pictorial data wherever possible. It is suggested that the analysis of spatial relationships between objects within a picture should work in the same way, subjects may perceive spatial relationships between such objects and therefore the possibility of using such relationships to search pictorial databases is proposed in this thesis.

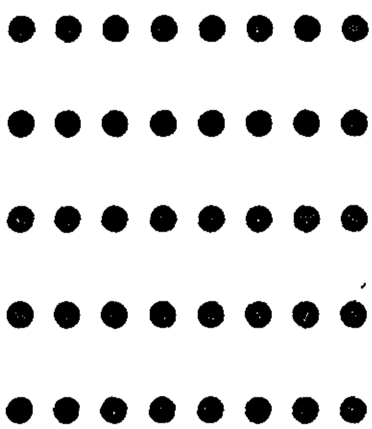
Many visual patterns can be explained in terms of Gestalt effects. For example the two shapes shown in figure 5.1 both appear to be squares irrespective of their actual size and pattern, this is because the brain 'orders' the information into a unit rather than perceiving each individual object as a separate entity.

FIGURE 5.1

SHAPES VIEWED AS SQUARES

The Gestalt psychologists devised a number of principles to describe how certain perceptions are more likely to occur that others. Proximity is the 'nearness' of elements to one another. Elements which appear close together are perceptually grouped together. For example in the case of figure 5.2 the pattern appears to show horizontal lines because of the closeness of the dots constituting the lines.

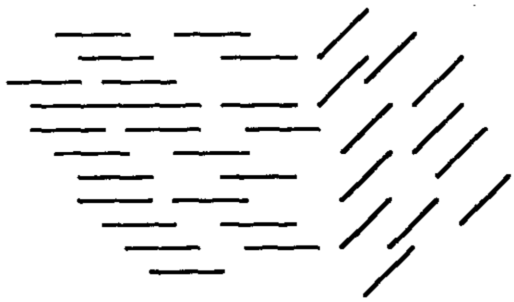
FIGURE 5.2



EXAMPLE OF GESTALT PROXIMITY

Another example of the Gestalt effect is similarity. Here elements of a picture which look similar are grouped together. Therefore in the case below all the horizontal lines appear to be grouped together, similarly the other lines at an angle to the horizontal also appear to have their own grouping.

FIGURE 5.3



EXAMPLE OF GESTALT SIMILARITY

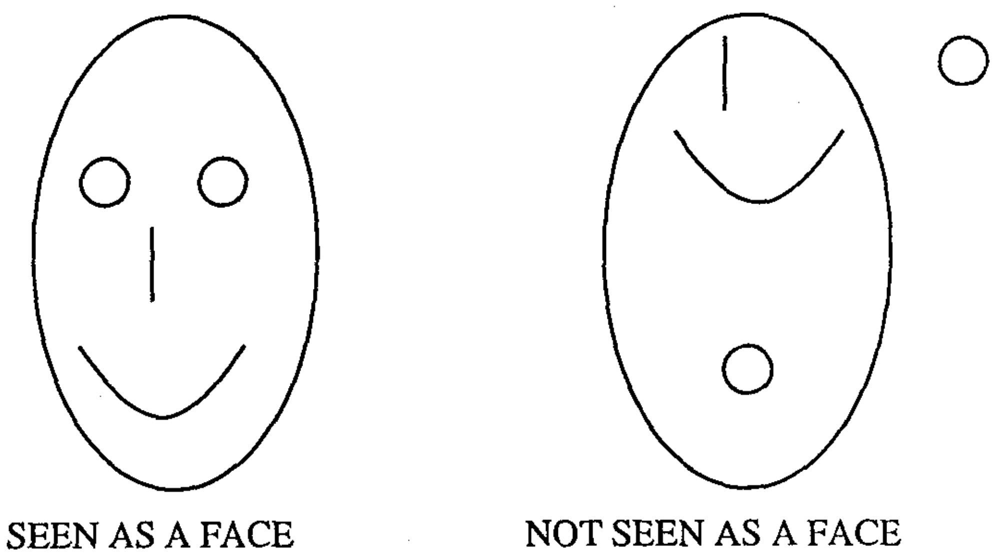
In attempting to understand why the Gestalt effects work, Marr's approach to vision (1976) emphasises that we should always consider what general assumptions about the world can be bought to bear on visual processing to constrain the range of possible interpretations possible of a particular image. Therefore these Gestalt principles may work because the assumptions that they embody about the world in general are correct. For example since the same kind of surfaces reflect and absorb light in the same kind of way then the different subregions of a single object are likely to appear similar. Thus any perceptual system which uses such assumptions to try and interpret natural images would usually achieve the correct solution to perceptual problems. For a designer encoding objects within a pictorial database system, the perception of groups of objects due to Gestalt principles needs to be taken into consideration in that groups of similar objects may need to be classified together within a picture, rather than encoding each similar object individually. This is because identical objects when grouped often appear to a viewer as a single object rather than as separate entities. For example a crowd scene in a picture may be better classified as 'people' rather than attempting to classify each individual person separately.

5.2.2.2 <u>The Perception of Spatial Relations</u>

It is suggested that one of the most important attributes in a picture are the spatial relations between objects contained within it. For the human visual system the perception of spatial properties and relations from a computational standpoint appear immediate and effortless and it far surpasses the capabilities of current artificial systems.

Visual perception requires the capacity to extract shape properties and spatial relations among objects and object parts. This capacity is fundamental to visual recognition since objects are often defined visually by abstract shape properties and spatial relations among their components. People are able to make judgment comparisons about objects of similar but nonidentical shapes. The question is how is this judgment made? An example is the diagram below which is seen as a face, primarily because of the shapes representing the eyes nose and mouth.Rearranging these items means that the picture is no longer seen as a face. Therefore the recognition of objects depends not only on the presence of certain features but also on their spatial arrangement.

FIGURE 5.4



OBJECTS FORMING PERCEIVED PATTERNS

Our capacity to establish spatial relations is manifested by the visual system from a very early age. Infants of 1-15 weeks respond to face like figures in preference to 'scrambled' face patterns (Fantz 1961). This suggests that people's ability to recognise spatial relationships is quickly learned during childhood and readily applied.

5.3 REPRESENTING SPATIAL INFORMATION

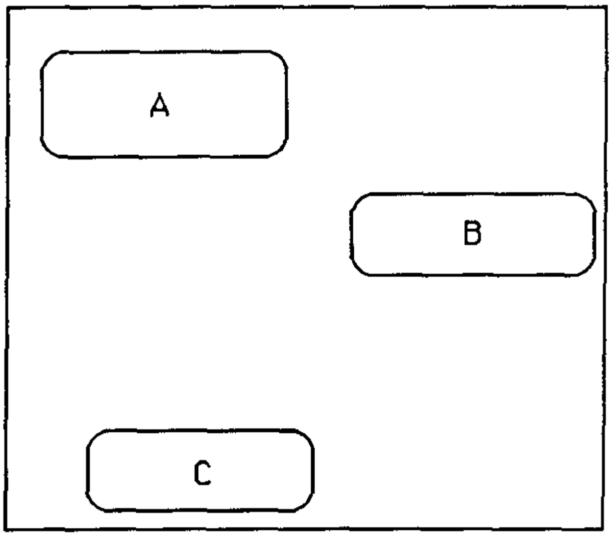
Many tasks require the use of knowledge that either pertains to or is derived from objects in either two-dimensional or three-dimensional physical space. Such knowledge is referred to as spatial knowledge (Srihari and Xiang 1989). Spatial knowledge is needed in tasks where either the input or goals involve spatial data. Examples include *spatial analysis* tasks such as image interpretation for computer vision and *spatial synthesis* tasks such as robot motion. Spatial knowledge can in general be either described or depicted. A descriptive representation would consist of such statements as left-of, right-of, above, below etc, here text is used. On the other hand a depictive representation displays the spatial relationships in the form of pictorial representations.

Spatial knowledge representations can thus be broadly subdivided into analogical (pictorial) and propositional (textual) representations. This broad categorisation has been arrived at not only by observing computer algorithms that use such knowledge but also from the theory of the left and right functionality of the human brain (Srihari and Xiang 1989). Analogical representations, where spatial information is displayed pictorially includes geometric features such and two-dimensional pictures and maps. Here there is a strong analogy between the actual objects within an image and their graphical or pictorial representations. Measurements such as distance can be directly taken from these representations in order to determine actual distances between objects within the corresponding image.

However under certain conditions there are problems with the analogical approach. In the case where there is no explicit boundary then adequate representation of object's spatial arrangements can presént a problem. Also where two-dimensional projections of three-dimensional objects occur then three-dimensional reasoning based upon these two-dimensional images may lead to inaccuracies in conclusions about the objects, this is further discussed in the next chapter. However within these limitations an analogical approach to pictorial analysis can be successfully applied. In general analogical representations are useful for encoding spatial configurations of objects and therefore visually representing the spatial relationships amongst them, but less useful for encoding semantic or logical knowledge about an object's spatial attributes.

Propositional representations are textual statements which have either true or false values. Propositional representations include predicate calculus, relational structures, programming languages and statements such as "object A is left of object B", "object B is above object C" etc. Pictorial databases have in the past represented spatial data primarily as propositional statements rather than analogical representations (see chapter 3). However as the spatial relationships between objects become more complex, it becomes more and more difficult to adequately represent spatial data propositionally. Another problem is that geometrical facts that are not recorded propositionally within a system are lost. For example where three objects are present within a picture (represented below by the letters A,B,C)

FIGURE 5.5



THREE OBJECTS WITHIN A PICTURE

and these objects are propositionally represented by the statements:-
A left-of B,
B above C,

If information about the spatial relationship between A and C is not explicitly stated then it will be lost when the relationships given are stored within a textual database. Information about how far A is to the left of B and how far B is above C is also lost on such textual encoding. This thesis however attempts to provide a more complete analysis of spatial data by allowing users to specify spatial relationships analogically using visual techniques as methods to locate and examine pictorial information. Here rather than describing spatial data in terms of textual relationships, numerical coordinates are used and these coordinates are used to identify spatial information. This analogical approach is examined in greater depth in the next chapter.

5.4 VISUAL MEMORY

5.4.1 Organisation of Objects and Spatial Relationships in Visual Memory

Mandler and Johnson (1976) have reasoned that there are at least three kinds of memory-relevant information a subject can extract from a picture which might be influenced by the nature and organisation of that picture:-

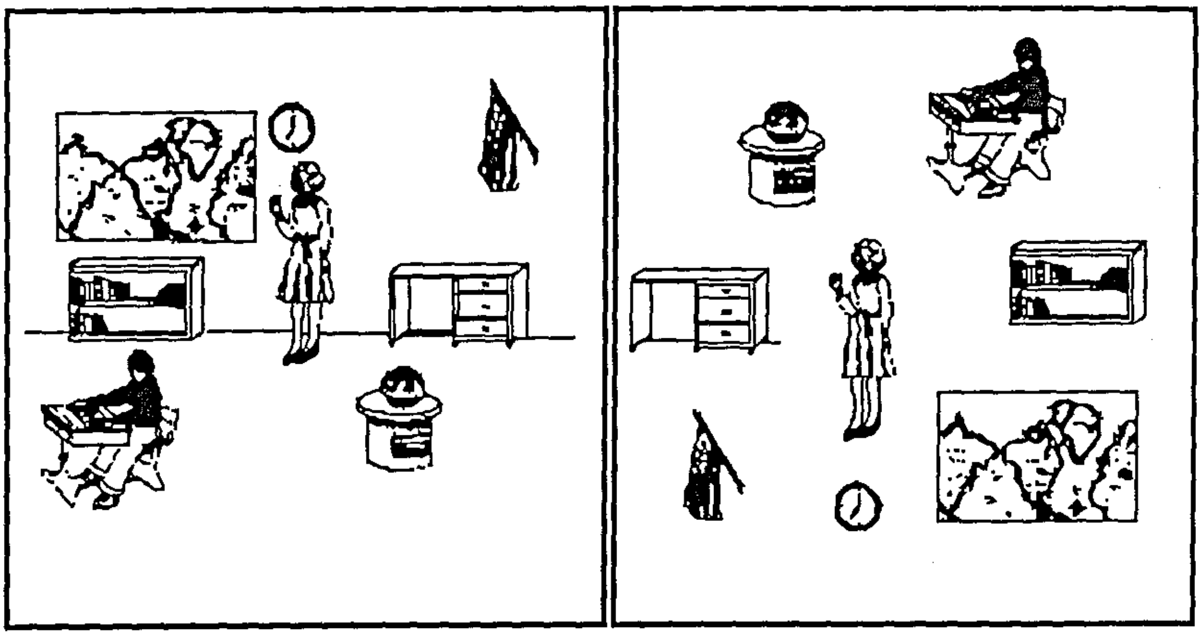
1) An overall description of the picture (ie its theme) identified from an inventory of the objects.

2) Descriptive information about the objects in the picture, for example details of their physical appearance, size and orientation.

3) The actual location of the objects as well as the spatial relationships between them.

In order to test this theory Mandler and Parker (1976) asked two groups of subjects to look at either organised or unorganised pictures for twenty seconds, after which the subjects were then questioned about their memory of the pictures. Both groups were found to have the same performance results in remembering detail, size and orientation of the objects within the picture in question, although organising the picture into a 'real-world' form (as shown by the left hand picture in figure 5.6) did help the groups remember the physical locations of the objects present. This was not only due to the fact that some objects in an organised scene were restricted (boys sitting at desks do not float in mid-air!), because the horizontal locations of objects were also very accurately positioned by the subjects as well.

FIGURE 5.6



ORGANISED PICTURE UNORGANISED PICTURE

ORGANISED AND UNORGANISED PICTURES

The degree to which spatial information was remembered was tested by asking subjects to study pictures after which chosen objects from the pictures were given to them. Subjects were then asked to place these objects within a blank picture frame in the spatial locations in which they remembered them. When the same subjects were later tested again there was a tendency by the subjects to make unorganised pictures conform to organised scenes. This suggested a loss of memory for the specific pictures and the use of knowledge of the given objects in the real-world to guide construction at the time of recall. However the organisation of the pictures appeared to have little effect on memory for the descriptive characteristics of objects. In contrast to this, organisation of the pictures had a marked effect on reconstruction of spatial information within the scenes. It was therefore suggested by Mandler and Parker that real-world organisations play two roles in visual memory:-

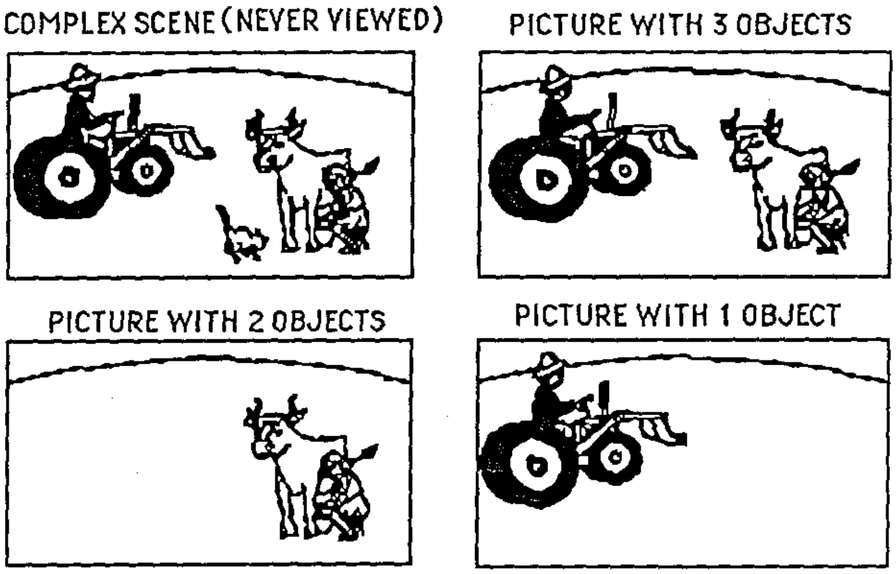
The first role is the *encoding* of visual information. The vertical dimension of pictures contains most of the invariant real-world knowledge about spatial locations, and therefore in organised pictures vertical placement was found to be more accurate than horizontal placement. In unorganised pictures vertical information is apt to conflict with real-world organisations of objects, whereas horizontal information has approximately the same degree of indeterminacy as in organised pictures. Thus this should explain the fact that subjects vertical placement was considerably less accurate than horizontal placement in unorganised pictures when subjects placed objects on a blank page. Since performance on both types of information was better in organised pictures than unorganised pictures this further suggests that spatial relationships in general are better encoded into a subject's brain if they fit a real-world schema.

The second role played by real-world organisations is one affecting *memory* for spatial relationships and may be either supporting or distorting. The real-world organisations carrying primarily vertical information supports memory for vertical relationships in organised pictures. This schematic support consists of real-world knowledge of the way things are supposed to be arranged. The systematic distortion on the vertical dimension in unorganised pictures also testifies to the constructive nature of memory.

People often use locational information to help organise objects for recall and it appears that a large quantity of locational information is automatically coded into long term memory in the sense that active processing of this information is not required (Mandler, Seegmiller and Day 1977). Therefore it would appear that positional memory about a scene persists long after the actual knowledge of the objects within it becomes hazy. Even when subjects are instructed to remember other physical attributes of objects within a picture such as colour or shape, the memory for these attributes is not as efficient as memory of the spatial arrangements of the objects.

Pezdec (1978) carried out an experiment to see whether subjects picture memory retains pictures in terms of the general theme of the scenes viewed, or alternatively whether the individual picture was remembered. This was tested by having different numbers of objects in a picture such that subjects viewed a series of pictures that contained one, two or three of the objects within the complex picture, but never viewed the complete picture. The three pictures viewed and the complex scene which was never viewed are shown in figure 5.7.

FIGURE 5.7



COMPLEX SCENE AND PICTURES CONTAINING COMBINATIONS OF OBJECTS SEEN BY SUBJECTS

The fact that many of the subjects thought that they had seen the complex scene when in fact they had only seen objects within it in separate pictures indicated that memory processes pictures for theme, and subjects do not always remember the exact visual details contained therein, in effect there is partial recognition. It therefore appears that pictorial memory needs the ability to use organisation to construct meaningful representations of integrated visual ideas in order to achieve maximum effect. Unfortunately since theme rather than specific attributes of objects are used as a means by which pictures are remembered this makes the rigorous analysis of information within pictures more difficult. This is because different subjects may perceive the same picture to be represented by different themes. Therefore identifying pictures based upon descriptions of such themes would therefore lead to confusion.

5.4.2 Duration of Picture Memory

Visual memory is in general exceptionally good. The two most important aspects of memory are its large capacity and long duration. Subjects when asked can usually remember scenes they have only seen once and can also recall pictorial features and places during their childhood. Although they are unable to describe objects in their childhood scenes with any great detail, the overall recognition of these scenes are exceptional (Shepard 1967).

In order to verify this, Shepard had subjects study 612 coloured pictures one by one, these subjects were then later tested on 68 pairs of pictures, one of the pictures in the pair had been present in the 612 pictures shown to them earlier, the other had not. The subjects were then asked to identify which of the pictures from the pair they had seen before. The result was 97% correct response in identifying the correct picture. Even after seven days this figure only dropped to 87%. This shows that people can recognise given pictures over a considerable period of time and although if asked for exact details of the pictures their recall is distinctly hazy, recognition of the pictures concerned is excellent. Pictures would seem therefore to afford a great deal of easily retained information to the observer in a short time.

Although people can remember objects and scenes given enough time to do so, an important point is whether given a much shorter time to observe an object, is it still possible to extract information from a scene and if so is the amount of information that can be extracted directly proportional to the time the scene is viewed.

Potter and Levy (1969) investigated this and found that a time span of approximately two seconds was required for the accurate recognition of an object. If given less than two seconds to observe a scene, subjects were still able to recognise that scene later, although detecting whether the object in question was in the picture usually needed at least two seconds for accurate recognition. Thus Potter and Levy demonstrated that memory recognition for pictures is a function of the exposure duration of the stimuli presented.

These conclusions would imply that pictorial database management systems should attempt to take account of picture themes, or at the very least objects that would be expected to be found in pictures of a given theme. This would enable users searching for pictures to locate them based upon the objects in the picture that they are searching for. Obviously this concept can not be used in all circumstances, however there are many cases in which a detailed identification of many of the smaller objects within the picture can enable the picture to be more readily located from a pictorial database.

5.4.3 Spatial Information Its Coding and Recall

The use of space as a means of organising thoughts or ideas dates at least back to the formation of the Method of Loci during the fifth century B.C. More recently there has been an attempt to give abstract objects of computer filing systems such as commands and files physical locations as well (for example the Apple Macintosh interface). Many new systems employ desktop metaphors in which an object can be given a location instead of, or in addition to a name, through its placement in a two-dimensional space on a computer screen (Jones and Dumais 1986).

Jones and Dumais have suggested that the fact that some locational information can be recalled even though the content of that information is often forgotten shows that spatial information is stored in part at least in long term memory. For example they found that people remember the location of articles in newspapers long after the exact nature of the article has been forgotten. Other kinds of information about the article such as colour, font and length may also be forgotten, yet the spatial layout of such information whether pictorial or textual seems to be retained considerably longer.

However Jones and Dumais's experiments have suggested that the efficiency of a spatially oriented approach to object reference in a computing system may be limited. They found that that subjects remembrance of locational information was worse than for name-only information, and that the combination of symbolic and spatial information produced only a modest improvement. They further suggested that additional objects when added to an information space may actually crowd and interfere with memory for object recognition until familiarity with the additional objects has been obtained. Unfortunately their experiment used textual news articles as the database objects rather than pictorial information which may have strongly influenced their conclusions.

In general it has been argued (Schulman 1973) that memory for events depends upon the ability to recall the attributes of those events. Indeed Underwood (1969) goes so far as to propose that memory for an event consists of a collection of attributes. Schulman considers that events can have two types of attributes, those that can be considered apart from the event and those that are inseparable from it. He suggests that locational information appears to be an attribute of the latter sort and that location is an attribute that is routinely encoded. Schulman found that the greater a subject's certainty of recognition then the greater the accuracy of locational recall. He concluded "that a object's location resulted in poorer locational recall and poorer memory of recognition when the subject received instruction to attend to the object's spatial location rather than where the objects were 'incidentally' learned". However as with Jones and Dumais, Schulman used words rather than pictures as the objects within the experimental spatial arrays. Mandler, Seegmiller and Day (1977) extended this work by addressing the question of whether locations of objects are automatically encoded in memory in the sense of being remembered even when a person does not consciously try to remember them, they also analysed whether a deliberate attempt to remember the locations of objects interfered with memory for the objects themselves. Sixteen small toys were placed in a wooden matrix of 36 locations where no more than two belonged to the same conceptual category (animals, vehicles etc). Mandler, Seegmiller and Day suggested that many of the incidental conditions that had been used by other researchers in examining spatial information (such as Schulman above) where in fact not incidental. They postulated that when subjects were asked to remember the nature of the objects and were then instead tested on the spatial location of the objects, that the subjects may have used spatial techniques as a mnemonic device by which to remember the nature of the objects, thereby making rigorous analysis of the incidental spatial information associated with the objects impossible.

that additional objects when added to an information space may actually crowd and interfere with memory for object recognition until familiarity with the additional objects has been obtained. Unfortunately their experiment used textual news articles as the database objects rather than pictorial information which may have strongly influenced their conclusions.

In general it has been argued (Schulman 1973) that memory for events depends upon the ability to recall the attributes of those events. Indeed Underwood (1969) goes so far as to propose that memory for an event consists of a collection of attributes. Schulman considers that events can have two types of attributes, those that can be considered apart from the event and those that are inseparable from it. He suggests that locational information appears to be an attribute of the latter sort and that location is an attribute that is routinely encoded. Schulman found that the greater a subject's certainty of recognition then the greater the accuracy of locational recall. He concluded "that a object's location resulted in poorer locational recall and poorer memory of recognition when the subject received instruction to attend to the object's spatial location rather than where the objects were 'incidentally' learned". However as with Jones and Dumais, Schulman used words rather than pictures as the objects within the experimental spatial arrays. Mandler, Seegmiller and Day (1977) extended this work by addressing the question of whether locations of objects are automatically encoded in memory in the sense of being remembered even when a person does not consciously try to remember them, they also analysed whether a deliberate attempt to remember the locations of objects interfered with memory for the objects themselves. Sixteen small toys were placed in a wooden matrix of 36 locations where no more than two belonged to the same conceptual category (animals, vehicles etc). Mandler, Seegmiller and Day suggested that many of the incidental conditions that had been used by other researchers in examining spatial information (such as Schulman above) where in fact not incidental. They postulated that when subjects were asked to remember the nature of the objects and were then instead tested on the spatial location of the objects, that the subjects may have used spatial techniques as a mnemonic device by which to remember the nature of the objects, thereby making rigorous analysis of the incidental spatial information associated with the objects impossible.

Mandler, Seegmiller and Day conducted an experiment wherein to produce a truly incidental condition, subjects were asked to price individual objects (small toys) in order to estimate the overall price cost of the object array. They postulated that locational information would not be encoded in this case. Two other conditions were also examined, the intentional condition in which the subjects were told to remember both the nature of the objects and their spatial locations, and the standard incidental condition in which subjects were told only to remember the nature of the objects present.

The results from these experiments appeared to demonstrate that a great deal of spatial information is available for retrieval without attention having been directed to it. Within the first experiment the loss in location information retrieved in the true incidental learning condition (in which the objects were priced) compared to the intentional learning condition was only 18%.

The experiments showed that there was no cost to object recall when subjects were required to remember locations as well. Analysis of the results appeared to show that processing of an object's location is different from the processing of other object attributes such as colour. In processing of the colour attribute there is a trade-off in which attention to the attribute of colour results in a lesser degree in the processing of the nature of the objects themselves. Thus the higher the attention paid to colour the less is the attention paid to the nature of the objects. In the case of the truly incidental spatial condition in which the objects were priced, the loss in the retrievability of locational information was no worse than it had been when the objects alone were recalled. Therefore Mandler, Seegmiller and Day postulated that conscious processing does not seem to be required for spatial information to be encoded into long-term storage, they suggested that locational information is encoded into memory irrespective of the attention paid to other physical attributes of objects.

Hasher and Zacks (1979) extending this work and proposed that spatial, temporal and frequency information are all processed automatically. They defined an automatic operation as one that drains minimal energy from our limited-capacity attentional mechanism and that this does not interfere with other ongoing cognitive activity. As a result, until recently, the claim for the automaticity of memorisation of spatial information about the objects within a picture seems to have had solid support (Hasher and Zacks 1979).

5.4.4 Coding of Spatial Location, An Automatic Process?

Naveh-Benjamin (1987) examined Hasher and Zacks work and found that the results of some of the tests of criteria for automaticity were rather ambiguous, whereas other criteria had not been studied at all. In examining Mandler, Seegmiller and Day's work he also concluded that a drop of 18% in locational information was a significant difference between the results of intentional learning and true incidental learning and that Mandler, Seegmiller and Day's claim for only a small loss in locational information could not be substantiated.
Using a 6 by 6 matrix in which 20 of the 36 locations contained items to be recalled, Naveh-Benjamin examined the coding of spatial location. However unlike Mandler, Seegmiller and Day's work in which real toys were used, Naveh-Benjamin used drawings of objects in a paper matrix to examine spatial information. The items within the matrix were chosen to be easily memorable and again no more than two belonged to the same conceptual category. In addition to this no more than four objects occurred in any row or column of the matrix and no more than five drawings of objects were present in any quadrant. Having constructed this matrix Naveh-Benjamin on testing subjects recall found that:-

1) The encoding of spatial information was influenced by the intention to learn. Subjects who learn the information intentionally recalled more accurately both the absolute positioning of the objects and their spatial relations when compared with subjects who incidentally learn such information. In addition the encoding of spatial location information was found to be a function of the subject's age. Older subjects tend to do less well than younger subjects in the recall of spatial information, however under intentional instruction the difference between the two age groups was much smaller.

2) The encoding of spatial location information was influenced by competing task demands. As the load of competing tasks is increased then the accuracy of spatial

location judgments declines. This was shown by having two groups carrying out concurrent tasks, one group of which had a considerably harder task to do. The group with the harder concurrent task was found to have a poorer recall memory for location than the group who studied location with an easier concurrent task.

3) The encoding of spatial location information was influenced by practise. Training subjects in the techniques of spatial learning results in a significant improvement in performance for memorising absolute spatial locational information. Again this was determined by examining recall of objects within the paper matrix.

4) The encoding of spatial location was influenced by individual differences in ability and motivation (attentional capacity) of the subject. Here two student classes of differing ability were tested and the results compared. The two classes were distinguished by previous entrance exam results. The class with the higher exam results were found to have a greater recall of objects within the matrix.

All the conclusions above were determined using recall of the items within the matrix. This recall task was quite complex in that the subjects were asked to reconstruct the whole matrix of objects by placing objects in their original positions. Naveh-Benjamin postulated that such a task might involve several processes at the time of retrieval such as searching memory for location information, retrieving several items first, and then reconstructing the locations for the other items. If it were possible that the encoding of spatial location was automatic then the use of such a complex sequence of tasks which combine to produce the recall task might obscure it.

In order to verify the four conclusions above, Naveh-Benjamin carried out further experiments (Naveh-Benjamin 1988) in which a simple recognition (rather than recall) test of spatial location was carried out in which subjects had to judge which of the objects from the matrix were presented in their original spatial location. Naveh-Benjamin postulated that this would reestablish the encoding context, minimise the effects of active search on the retrieval task and probably require less strategic processing.

The results found were in agreement with the four conclusions above in that:-
1) Subjects who learn the information intentionally more accurately recognise the spatial locations of the objects when compared with subjects who incidentally learn the information (subjects who intentionally recognised the spatial locations on average remembered 12.9% of the information more accurately).

2) Older subjects do not recognise spatial location as well as younger subjects. Therefore spatial location information is a function of a subject's age.

3) Applying a short training regime results in a significant improvement in recognition. Therefore the encoding of spatial information is influenced by practise.

4)Students from more selective groups were more accurate in recognising spatial location information. Therefore the encoding of spatial location information is influenced by individual differences.

Therefore these results verify those conclusions postulated by Naveh-Benjamin from his earlier recall experiments. These combinations of results imply that neither absolute nor relative locational information is automatically encoded since intention would seem to be a prerequisite for accurate visual memory. However there is the possibility that subjects memory encoding of spatial location consists of a continuum linking automatic and non-automatic processes. If such is the case then these results imply that spatial location coding is not at the extreme automatic end of this continuum. Yet the results above have shown that there is considerable spatial encoding in the brain since subjects examined for incidental spatial memory still recall and can recognise many objects within a picture. Naveh-Benjamin suggests that the processes used to recognise objects within pictures include the initial encoding of the objects present into the brain and later retrieval of the objects from the brain for comparisons of their spatial location with viewed pictures. If one of these processes is not automatic there will be a difference between experimental results in analysing the automaticity of spatial information. Ideally what is required are attempts to further partition all the processes involved in memorising spatial locational information in order to assess their characteristics and to determine those which are automatic.

It has also been suggested (Sanders et al 1987) that a somewhat more lenient criteria for automaticity could be used instead of Hasher and Zacks' definition in which automatic operations are defined as "those that drain minimal energy from the limited-capacity attention mechanism and do not interfere with ongoing cognitive activity". According to Sanders "automatic processing should be capable of producing an above-chance output while requiring no more than minimal resources and little or no conscious awareness". Using this definition of automaticity, then the concept of spatial location would be considerably more automatic.

To conclude it would appear that spatial location is not fully automatic, though where it appears on a continuum between non-automaticity and automatic processes depends very much on how automaticity is defined. In order to resolve this problem the various processes that constitute the recall and remembrance of locational information need to be further subdivided in order to determine which processes are themselves automatic and which processes are not.

The logical extreme of spatial information systems is the spatial database management system in which pictures are placed in a simulated three-dimensional space representing a setting familiar to the user (discussed in chapter 4). Within these systems it is not enough to know what is being looked for but also where within the three-dimensional space it is located. One important measure of performance within such systems is the accuracy with which the user can specify or find an object. The accuracy of object reference clearly depends upon the user's ability to connect an object's internal contents and its external positional representation (what it consists of and where it is). One of the important questions here is how do users memories for an objects location compare with their memories for that object's name, and how does this vary with the size of the database? Results suggest that users do not automatically remember the spatial information required to navigate around such a system but that there is considerably more than an incidental memory of the spatial information required for navigation. They further suggest that intention to remember how to navigate around such a system and practise in the use of these systems provide better results in locating information.

It is suggested that the conclusions about spatial information drawn by Naveh-Benjamin's work can also be applied to the analysis of pictures. These conclusions suggest that spatial information is remembered in part at least even when conscious attention has not been paid to the spatial location of the objects within a picture, although paying attention to the positions of the objects enables users to more accurately recall the spatial information associated with the objects at a later date. This implies that the user's ability to remember spatial information should be a suitable technique for locating pictures from a pictorial database. Again Naveh-Benjamin's work suggests that such an ability is improved with practise.

In many computer systems we have the freedom to use whichever of many possible retrieval cues (shape, size, location, name, orientation, length, width, etc) which are most likely to lead to success in locating objects within pictures stored in a pictorial database. We need not be constrained by the same constraints that apply in the real world. In the real world to find a book we first need to go to the bookshelf irrespective of the colour or size of the book. However in the case of a computer system, attributes such as colour or location may also be used as a means by which a book is found from within a pictorial database. Therefore in a database management system designed for the retrieval of pictures from a pictorial database it is important that object attributes are included in order to enable the user to locate pictures based upon the nature and values of these attributes.

5.5 MENTAL REPRESENTATIONS OF SPATIAL RELATIONS

Theories of the mental representations of the spatial relations among objects and how they are encoded in memory, can be distinguished in several different ways. Two of the most important characteristics are the structure and the contents of the mental representations. Here three classes of theories are explained (McNamara 1986).

5.5.1 Non-Hierarchical Theories

According to non-hierarchical theories, spatial relations among objects are mentally represented in propositional networks or in analogue representations with continuously varying properties. The important claim of this theory is that there is no hierarchical structure to the representation, everything is represented at the same level. An example of these theories has been proposed by Byrne (1979). According to Byrne, memory for an environment can be viewed as a network that preserves topological connectedness (the order of locations and turns) but not two-dimensional orientation and distance information (as it would on a map). Locations in a layout correspond to nodes in the network and paths between locations correspond to links between the nodes. He argues that this theory is supported by the fact that estimates of urban distance increases with the number of turns and that the estimate of angles between urban roads tends to be distorted towards 90 degrees regardless of the actual angle. The larger the number of turns then the longer to judge the distance which gives a subject the impression of a greater distance. Angular statistics are not recorded in spatial memory so tend towards 90 degrees.

5.5.2 Hierarchical Theories

According to hierarchical theories different "regions" of an environment are stored in different branches of a graph-theoretic tree. The mental representation is organised such that increasingly more detailed spatial knowledge is given at lower and lower levels of the hierarchy. Hierarchical theories can be divided into two classes of sub-theory depending on the kinds of spatial relations encoded in memory.

5.5.2.1 Strongly Hierarchical Theories

These theories propose that spatial relations are inferred from higher order spatial knowledge instead of between locations in different branches of the hierarchy. With this theory the spatial relations between two towns in the same county would be not stored in memory but inferred from knowledge of the relative positions of the towns within the county. This class of theories represents an extreme position in the trade-off between storage and computation.

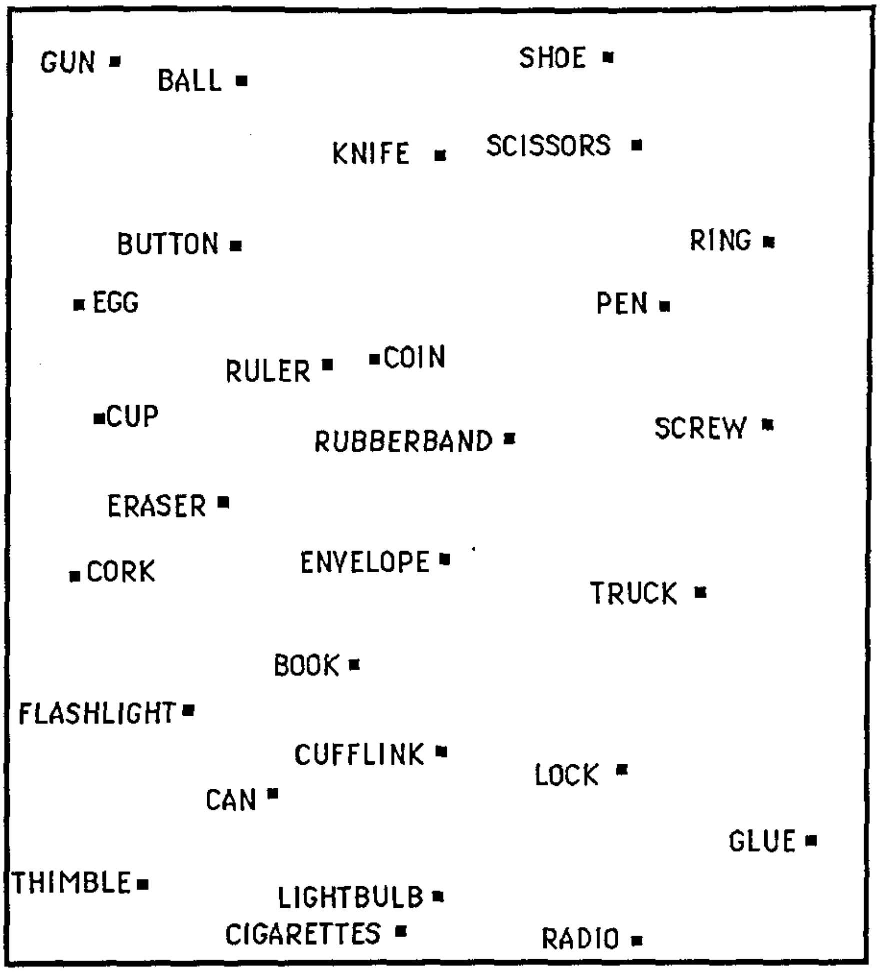
5.5.2.2 Partially Hierarchical Theories

These allow spatial relations to be encoded between locations in different regions of the environment. This is less efficient than strongly hierarchical theories in terms of storage of spatial knowledge however there is increased speed of spatial calculations.

In order to test these three theories subjects were shown a spatial layout and then participated in three tasks:- recognition of object names, judgment of the directions of objects relative to each other and estimation of the distance apart of the objects. Subjects also received tests of verbal and spatial ability. The results from all three tasks and all groups tested tended to support partially hierarchical theories. The major reason for this was the finding that locations in the same region primed each other more than locations in different regions. This indicates that locations in the same region were 'closer' in subjects memories than locations in different regions irrespective of the actual distance apart. This is consistent with partially hierarchical theories in which different regions are stored in different branches of a partially hierarchical mental representation.

McNamara, Hardy and Hirtle (1989) carried out further tests on the structure of spatial relations between objects by using a spatial layout of 28 real objects within a boundary of 6.1m by 6.7m. These objects were then viewed from above by subjects participating in the experiment.

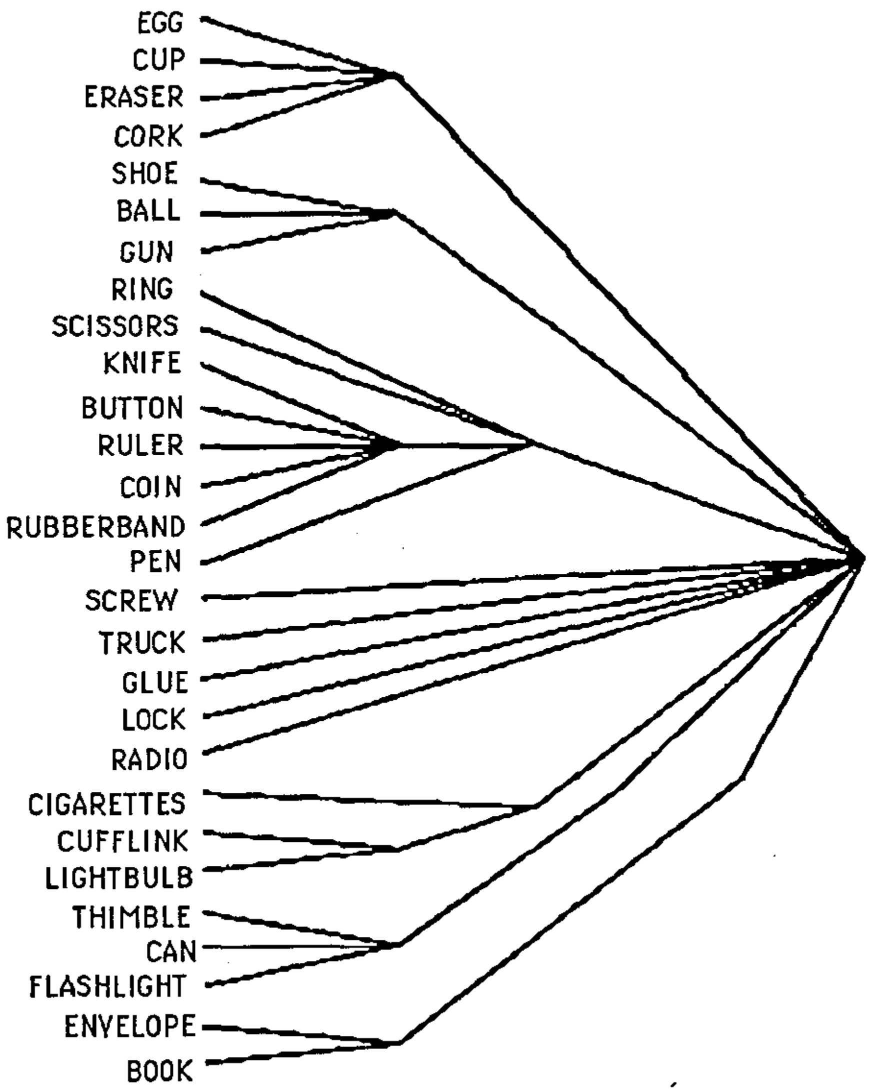
FIGURE 5.8



OBJECTS PRESENT WITHIN A BOUNDARY

In figure 5.8 the names of the objects are shown, however in the experiment the actual objects were used. This experiment examined item recognition, free and cued recall and distance estimation among the objects present within the spatial layout. Experimentally this entailed subjects memorising a set of items and then being asked to recall the items. This procedure was carried out several times and the order in which the items were recalled by the subjects was noted. Recall protocols from the results were submitted to the ordered-tree algorithm, producing hierarchical trees consistent with the internal organisations of individual subject's recall protocols. McNamara, Hardy and Hirtle found that subjects tended to group the same items together on recall. Within the ordered-tree algorithm, items that were recalled together were clustered together in the same sub-tree, this is shown in the diagram below. These trees modelled the organisations of subjects memories in that different sub-trees corresponded to different subjective regions in a subject's spatial memory for the positions of the objects. The appearance of these trees was several levels deep indicating that objects appeared to be grouped into several hierarchical levels. A hierarchical tree for one subject is shown below.

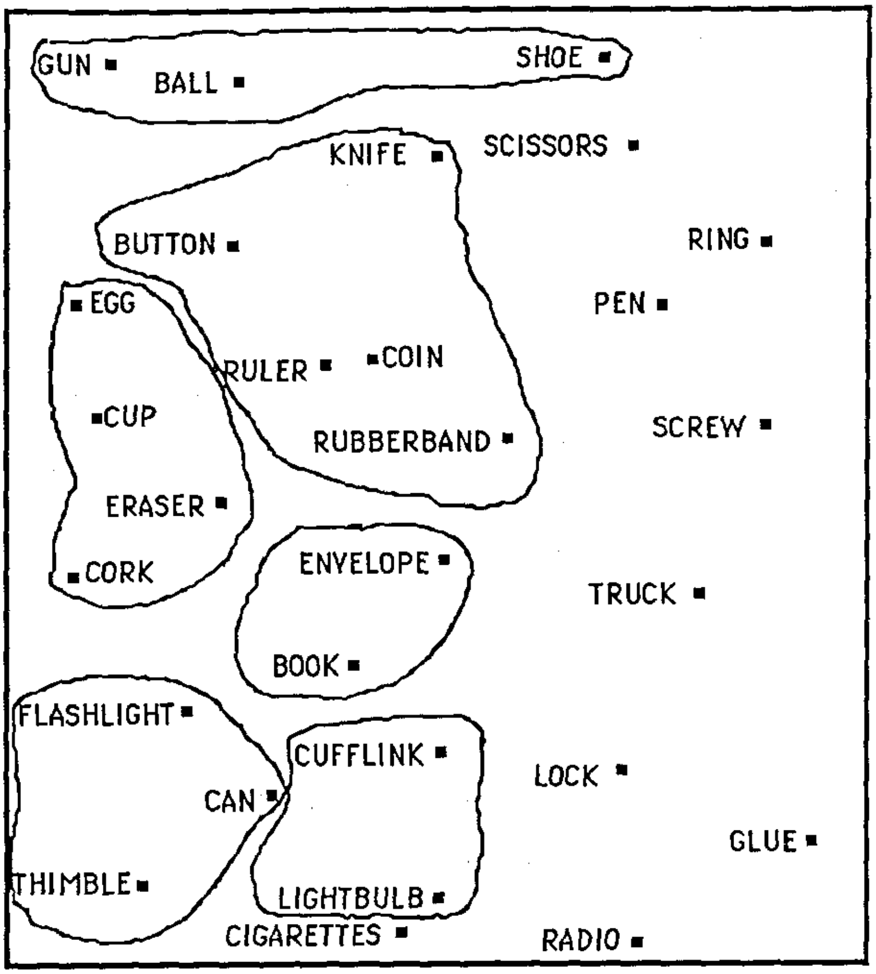
FIGURE 5.9



HIERARCHICAL TREE MODELLING SUBJECTS MEMORY

Thus at the lowest level of the hierarchy the objects were grouped into the organisations shown in figure 5.10.

FIGURE 5.10



SUBJECT GROUPING OF OBJECTS IN THE SPATIAL LAYOUT

The increase in time for item recognition and distance estimation showed the psychological distance between remembered locations was larger when those locations were in different subjective regions (sub-trees) than when those locations were in the same subjective region. Practically this means that when an item is remembered from a spatial layout, if a subject is queried about another item and the subject holds that second item in the same sub-tree in his/her memory, then this second item will be recalled faster than if the item is stored in a different sub-tree. Items within the same sub-tree are nearly always close to each other within the spatial layout, as shown in figure 5.10.

The tree produced shows that objects that might be considered to belong to the same semantic class such as book, envelope and ruler were not grouped together in memory. Objects appeared to be grouped based upon their spatial proximity rather than by similar semantic classes. Thus objects near to each other were grouped together rather than objects that were semantically similar but farther apart. This suggests that in analysing pictures people group objects in memory that are close together in pictures forming mental spatial relationships from such objects. Therefore it is suggested that such groupings of objects might be used in the analysis of pictures contained in a pictorial database.

Additionally within this experiment items stored in a different sub-tree took longer to recognise the larger the physical distance from an item in a different sub-tree. It was also found that subjects underestimated the distances apart of objects stored within a sub-tree since they grouped these objects together in memory. These findings argue that spatial memory has a hierarchical component, even when physical and perceptual boundaries are absent. However this research gives no indication about the form of spatial representations (ie whether they are propositional or imaginal) but it does constrain the structure of these representations. Such an experiment also suggests that given a picture in which there is a random arrangement of physically varied objects, objects are grouped spatially within the picture rather than semantically. Thus since users appear to group things spatially within a picture, a spatial technique of search in which objects are located (in part at least) by their spatial location should be a reasonable technique in finding pictures from a pictorial database.

Norman (1983) points out that the degree to which the user's mental model of a system matches the system designer's conceptual model of the same system, predicts the user's ability to use the system. Thus in a pictorial database where a designer extracts information from a picture for later comparison with the user's query, then where possible aid should be given in order to help the user understand the designer's mental model of the picture from which information has been extracted. There will be a poor understanding of the model if the domain is very complex and remote from the user's normal experience. Thus the aim should be to provide a system model which provides the inferential power of the conceptual model and which uses a language that the user is familiar with so that they can readily understand it. One such technique that can be used is that of direct manipulation in which the users place representations of objects on a screen representing the spatial arrangements of objects required within the pictorial database. This will be covered in greater detail in the next chapter.

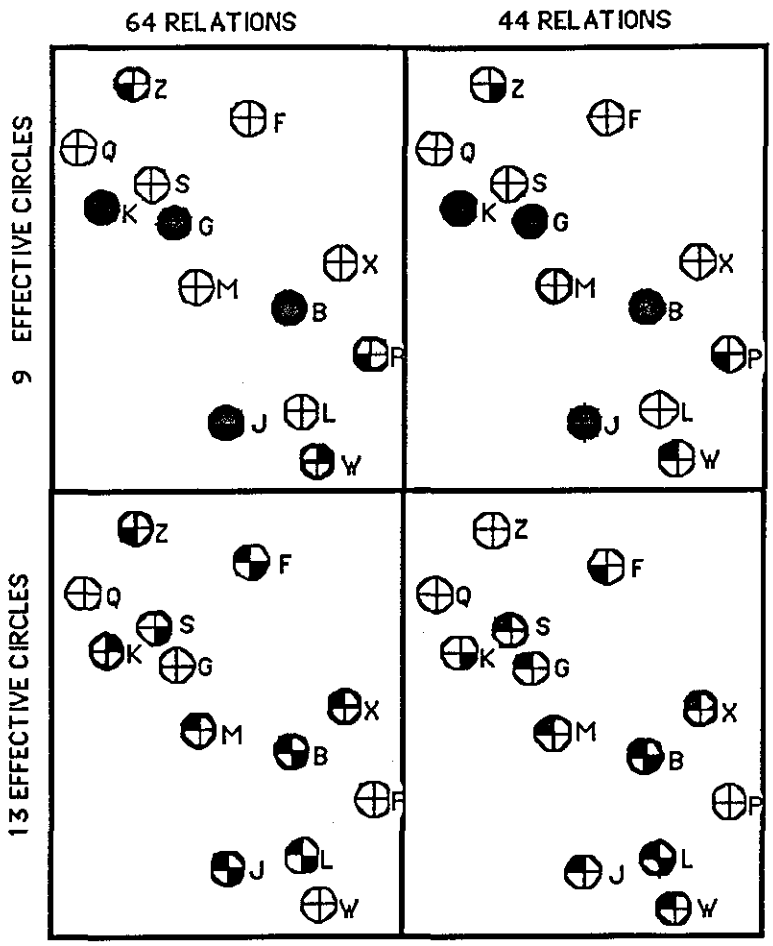
5.6 KNOWLEDGE OF SPATIAL RELATIONSHIPS

Another important aspect of spatial relationships is the mechanism underlying the way that spatial arrangements of similar or identical objects are remembered. This is particularly applicable to application areas having restricted domains in which objects present within a pictorial database of pictures can be represented by simple icons such as circles. Wilton and File (1975) conducted experiments investigating this.

In their first experiment, subjects examined an array of circles located in different positions on a background, after removal of the array subjects were then asked questions about the spatial relations between the circles. 13 circles were drawn each of 2cm in diameter and the position of the circles was placed randomly with the restriction that no circles overlapped or were horizontally or vertically aligned. Each circle was labelled with a different letter and divided into quadrants. Questions such as "If a straight line was drawn from the centre of circle X to the centre of a circle Y which quadrant of X would it pass through" were asked. The question was answered by specifying the appropriate quadrant of X: upper left, lower left, upper right or lower right. A subject's answer showed whether they knew how X and Y were spatially related. No questions were asked about circles whose line between them passed through the blackened quadrants. Due to these blackened quadrants questions could be asked about all 13 or only 9 of the circles, as indicated below. Another variable in the design was the number of questions that could be asked (44 or 64) about an array.

For example in the case of figure 5.11, the top left-hand array consists of 64 possible questions involving 9 circles (the other four circles being completely blackened). Thus these 64 questions consist of Z-7 (seven questions about a straight line drawn from Z to another circle) F-8, Q-7, S-8, M-8, X-7, P-6, L-7, and W-6.

FIGURE 5.11



ARRAY OF CIRCLES IN DIFFERENT POSITIONS

In the analysis of the arrays subjects could have tried to memorise the answers to the questions. This approach, a type of paired associate learning in which the question is the stimulus and the answer is the response, would be suitable if the subject had a large memory and no computational ability. However it was found that the subjects who should have learned 44 paired associates performed no better than subjects who were supposedly learning 64 paired associates between the circles. This non-effect of number of effective circles oń performance suggests that a subject learns the positions of the circles and that answers to questions are generated from the specification of a circle position.

This implies that our knowledge of spatial relationships is not knowledge drawn directly from memory of these spatial relationships; rather it is the consequence of our ability to manipulate stored

information to produce appropriate and novel responses. Therefore the positions of the objects are remembered and from these remembered positions, spatial relationships between these objects can be determined in response to a query.

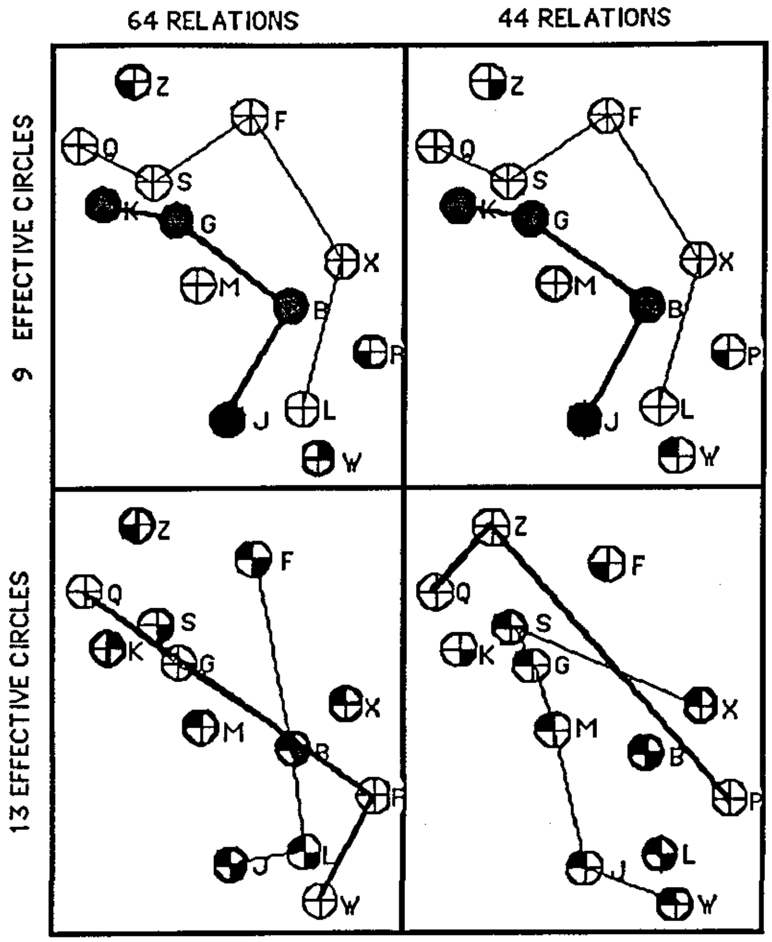
Thus there is a *distinction* between knowledge of spatial relationships between objects, and memory of their positions. Knowledge of spatial relationships is derived from other information rather than being directly contained in memory.

In order to investigate the structure of the stored information (ie the positions of the objects) and the computational procedures that are brought to bear on such information, similar experiments were carried out with an array of circles (Wilton and File 1975) and it was found that the ability to recall the arrangement of the circles' spatial layout depended on the structure of the array of the circles. This was tested by asking subjects to learn an array and then testing on their ability to recognise the array when only some of the circles were drawn on it. The results showed that when circles that were adjacent in the original array were selected for the test array, recognition was better than when circles were selected at random from the array, this suggests that only some of the relationships of the external array were used to specify the location of a circle and that these were relationships to nearby circles.

Subjects tended to group the circles into higher-order units and store these units rather than individual circles. Therefore the circles are grouped into patterns and positions of the circles were remembered by remembering the patterns they formed. These patterns occurred between nearby circles.

Therefore in the case of identical circles (within figure 5.11) the higher order units (in the form of lines connecting the circles) that were formed according to my own personal perception are shown in figure 5.12.

FIGURE 5.12



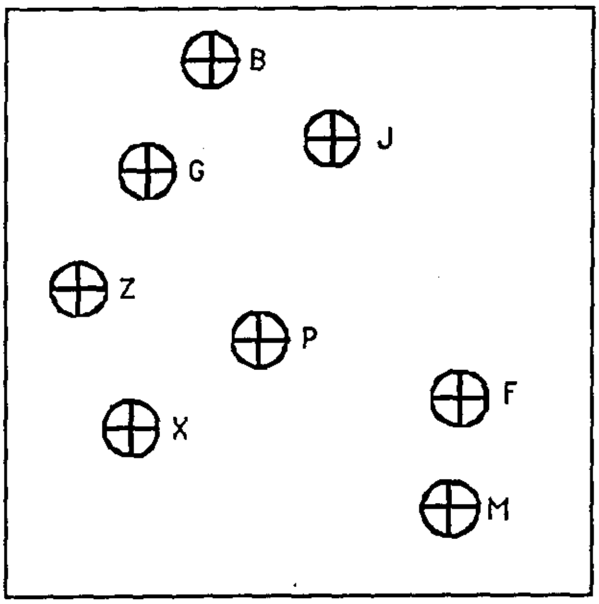
PERCEIVED HIGHER ORDER UNITS PRESENT

To store spatial locations in patterns may be advantageous in terms of memory load but it would appear to make the computation between selected circle locations more difficult. This is because there are now two levels of reference, memorising the pattern shapes between identical circles and secondly memorising how such patterns are themselves spatially arranged to each other.

Wilton and File (1977) extended this work by attempting to determine the nature of the memorised information by examining the order in which circles were recalled. In this experiment eight circles which had been split into quadrants were drawn. The circles were positioned randomly with the following constraint. For each of the circles in turn (circle B as the first example) then two test circles (circles J and F in this case) were placed to meet the following condition.

The angle between the centre of J, the centre of B and the horizontal = The angle between the centre of F, the centre of B and the vertical

FIGURE 5.13



EXAMPLE OF THE ARRAY USED

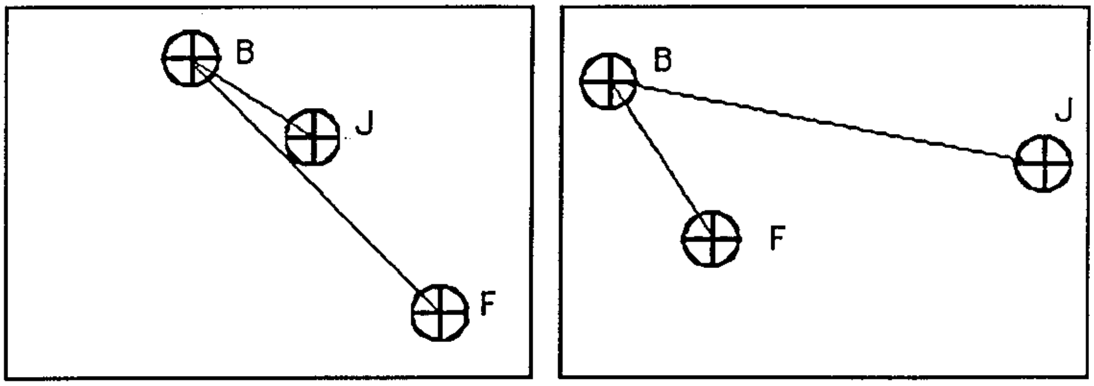
Questions were then asked such as if circle B had its vertical and horizontal lines extended, which circles would lie in the lower right hand quadrant of circle B. Here there would be four circles F,J,P,M. Of these only the order in which the two test circles (F and J) were remembered were important. It was found that when this procedure was carried out for various circles, the order in which the circles were recalled were for those circles nearer the top of the array before those nearer the bottom.

Thus the conclusion here is that:-

Given a specific object upon which to base the location of other objects, subjects recall other objects in a top-to-bottom fashion, rather than specifying the nearest objects first.

In the case of the two examples shown in figure 5.14 where the location is based upon circle B then circle J will be recalled before circle F irrespective of the distances between J and B.

FIGURE 5.14



CIRCLES REMEMBERED IN TOP-TO-BOTTOM FASHION

When the circles were shown to the user sequentially rather than simultaneously as in the case above and the same procedure carried out it was found that the order in which the two test circles were recalled varied with the order of input. This suggests that the subjects do not store the combination of serially presented circles in a spatial array. If a spatial array had been formed then the order in which the test circles were recalled should have been the same as in the case above and independent of the input order. Therefore it appears that under these circumstances a subject does not form an 'internal map' of the spatial layout of the circles. Therefore these results are compatible with the hypothesis that the positions of the circles are stored in a list of propositions rather than analogical representations (Pylyshyn 1973). However it is also possible that this list of propositions is converted into a analogical representation within the brain when enough circles have been seen.

Kosslyn (1987) explains this storage of spatial information by suggesting that spatial information can be stored in either a propositional or an analogue form. When the information is to be retrieved searches can be made in both forms, if the propositional form does not provide the answer then the analogue form will. An example of this theory is if somebody where to ask you the number of windows in the front of your house. If you had counted them before or if there was only a small number then the information is probably stored propositionally. However in other cases the information is stored analogically and a spatial representation is created from the knowledge you have of your house from which the windows can be counted.

5.6.1 Cognitive Maps

The knowledge of relative spatial layout is often called a cognitive map. A cognitive map is a form of memory which enables us to move from one location to another without getting lost. People use cognitive maps by recognising labels placed with the map, for example people remember routes by means of recognising objects along that route.

Baum and Jonides (1977) found that the time to judge distances within a cognitive map was directly proportional to the actual distances involved. Therefore subjects were using a picture of the path they were going to take. This was demonstrated by showing students the name of a building on their university campus and then giving them the name of two other buildings and asking them which was the nearest to the building in question. Baum and Jonides found a positive correlation between decision times and the distance ratio.Therefore people rely on two primary sources for the information in their cognitive maps: direct experience and cartographic information. Evans and Pezdek (1980) tested subjects knowledge of relative spatial positions for two kinds of items. One was a cartographic map of the United States of America and the other was the subject's college campus. The major difference here is that the subjects had seen the college campus from many different visual orientations but could only see the cartographic map from one orientation.

On the cartographic map, triads of towns were represented by dots, and the subjects had to determine whether the dots representing the town or cities were correctly positioned or had to be rotated in order to be correct. Evans and Pezdek found that the decision time for the rotation of the towns increased with the degrees of rotation the towns and cities were away from the correct positions. However using the campus buildings on the subject's university, the orientation of the buildings did not affect the decision times. Here cognitive map information had been acquired in an orientation free manner.

To conclude, what people remember about pictures depends largely on the purpose for which they observe them. However it would appear that certain characteristics of pictures have a degree of automatic encoding. The overall 'theme' of the picture as well as the spatial attributes of objects appear to be retained in memory long after the exact nature of the objects within a picture have been forgotten. Therefore it would seem reasonable to try to use both of these characteristics as search techniques in the analysis of pictures. However since different users interpret the 'theme' of the same picture in different ways, spatial positions and spatial arrangements of objects within a picture are suggested to be suitable attributes when attempting to locate pictorial information from a pictorial database.

Analysis of pictures therefore requires a detailed study of the nature of the objects present within a picture using techniques such as imagery and pattern matching, along with a study of how effectively the user is able to remember the spatial arrangements of such objects and the mechanisms for doing so.

5.7 PICTURE ANALYSIS

5.7.1 Eye Movement (Fixations and Saccades)

One of the most effective methods of understanding how pictures are processed and what objects people concentrate on and remember is to observe how people move their eyes when they look at pictures. Specifically there are two recognised types of eye movement, fixations and saccades (Yarbus 1967). A fixation is a time when the eye is stationary on a fixed point within a picture. A fixation can last between 200 milliseconds and 500 milliseconds and usually occupies between one and five degrees of the visual angle of view. In between fixations saccades occur in which the eyes' focus is changed to a new part of the picture. During saccades visual information is blurred across the retina and very little is recorded.

In order to observe what people are looking at while viewing a picture, Yarbus carried out an experiment in which a narrow beam of light was reflected off the surface of a subject's eye and the deviation of the beam produced measured. This was achieved either by the subject wearing a contact lens or reflecting the beam off the cornea (part of the eye) of the subject.

Yarbus noted that a very large proportion of the fixations were made up on areas that were highly informative containing information that details the nature of the objects observed. In scanning the picture of a human face, subjects were most likely to fixate on the eyes, mouth and nose, these areas being the most informative part of the human body giving a good indication as to how the person observed is feeling. However fixations are not necessarily drawn to area of greatest detail, neither are they frequently carried out on regions of extreme darkness or brightness. People look more at visual information that gives an overall impression of the picture, rather than specific objects within it. Yarbus also found that one of the factors that determined the scanning pattern followed by the subject's eye, was the subject's purpose in looking at the picture in the first place. Thus for example subjects scan objects in pictures differently if they are first asked to determine the age of objects in the picture and then the cost of objects present. Similarly in the case of shapes rather than real-world scenes Berlyne (1958) found that given two stimuli, subjects always focused on the less predictable of the two members, this is due to the fact that subjects are able to determine the nature of predictable shapes by forming high order patterns derived from the combinations of the elements of the picture.

Three types of eye movement have been distinguished by the situations in which they occur (Kahneman 1973). The first is *spontaneous looking*. This type of looking is governed by the novelty and complexity of visual stimuli. The second type of looking is *task-related* looking. The eye has sharp vision at its centre while peripheral vision is increasingly less distinct towards the outer boundary of the eye. The problem of where to look next is resolved through the interaction of task constraints and the visual environment. The third variety of eye movements involves looking as a function of the changing orientation of thought. Movements of the eyes appear to reflect transitions between modes of thought about the nature and description of the picture even though, in many cases, where a subject is looking can offer no new information.

Antes (1978) continued the work of Yarbus and Berlyne by looking at the duration of fixations and saccades when subjects see a picture for the first time. Initially the subjects made a brief fixation followed by a long saccade as if scanning the picture. After this, the fixations are longer and saccades shorter, first fixating on a subject and then moving to another subject nearby. In general eye movements appear to reflect what the brain is thinking about. This has been shown by a simple experiment in which Cooper (1974) told stories about certain objects. As the subjects listened to the story and heard the name of a specific object, their eyes automatically focussed on the object contained within a display in front of them. The actual story concerned solid objects such as lions and zebras. The subjects were not aware that their eye movements were being followed. It was also found that they fixated on pictures of things conceptually related to items in the story. On interviewing the subjects Cooper found that their eye movement appeared to be automatic without any conscious planning on the part of the subjects. Therefore as before with Yarbus's experiments, if users have a purpose in looking at a picture (whether instructed to in the case of Yarbus's experiments or unconsciously in the case of Cooper's experiments), then they focused more often on objects within the picture which were related to the query given, rather than other objects within a picture.

There is difficulty in applying Yarbus's, Berlyne's and Cooper's findings to the design of pictorial database systems since the objects or spatial arrangements required by the user from within the pictorial database are often difficult to predict. The designer may encode the most informative or distinctive features present within a picture however this encoding may not include the object or spatial arrangement of objects that the user requires. This suggests that where possible the information encoded from a picture should be determined by consultation with the user in order to determine the nature of the objects required. Having consulted with the user as to their requirements, the designer of the system can then encode the most appropriate information into the data management system.

5.7.2 Practical Picture Constraints

Biederman (1981) observed that to comprehend a pictures content requires not only an accurate inventory of the objects present, but also an understanding of the functional and spatial relations among them. Viewers must arrive at an overall representation that serves to integrate all the aspects of the picture. This representation or *Schema* is a mental structuring of data that embodies necessary real-world constraints on the content and organisation of the picture and thus contains expectations about what should be contained within the picture. Biederman found five important types of constraint imposed by pictorial organisation, support, interposition, probability, position and size.

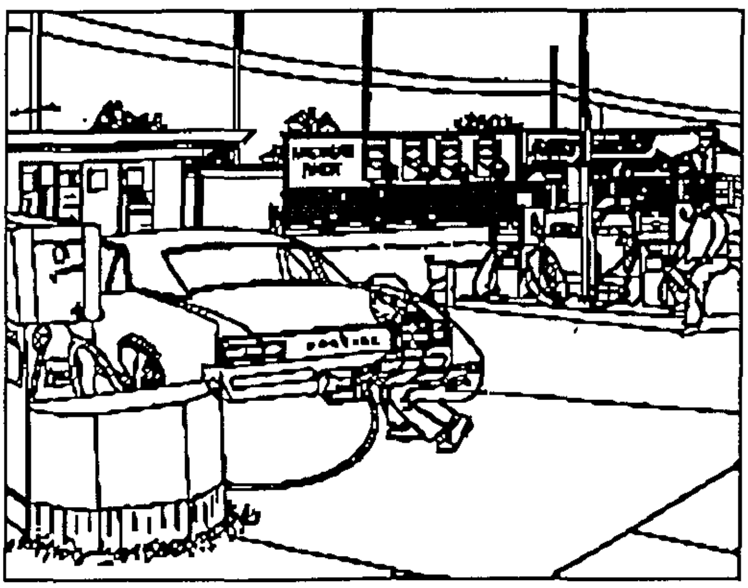
1) SUPPORT

Objects do not defy the laws of gravity. For example cars do not float in mid-air, therefore a flying car would break this constraint. However it is recognised that flying objects such as birds and planes are not restricted in this way.

2) INTERPOSITION

Objects can not be seen through other objects unless they are transparent, therefore being able to see a car through a man as in the example below would break this constraint.

FIGURE 5.15



PICTURE WHICH BREAKS THE CONSTRAINT OF INTERPOSITION

3) PROBABILITY

Certain objects are much more likely to appear in a scene than other objects. For example seeing a piano underwater is extremely unlikely.

4) POSITION

Seeing a lorry on the top of Mount Everest would break the constraint of position. This constraint is closely related to probability.

5) SIZE

This constraint applies to keeping the objects in question in relative size to each other. For example a man the size of an elf sitting on a persons shoulder would break this constraint.

FIGURE 5.16

PICTURE WHICH BREAKS THE CONSTRAINT OF SIZE

In figure 5.16 a miniature man appears to be present which breaks the constraints of size. Objects can break more than one constraint at any one time and the more constraints that are broken the easier the offending object is to see and identify. Each constraint that is violated within a picture helps the subject to more readily identify those constraint violations. Therefore pictures in which three pictorial constraints are broken are easier for a subject to identify than those where only one constraint is broken.

The implications of these conclusions are that subjects expect pictures to meet certain real-world constraints when viewing them. They therefore make certain assumptions about the spatial organisation and arrangements of objects within pictures. Thus it is suggested that in real-world scenes in which these constraints normally exist, these constraints can be used in the identification of pictorial information by visually depicting information.

5.8 SHAPE RECOGNITION

Another important aspect to picture study is the ability of recognise shapes, one of the most recent theories of which is Marr and Nishihara's.

5.8.1 Shape Recognition and Three-Dimensional Models

Marr and Nishihara (1978) have proposed that "the shape recognition process defines a coordinate system that is centred on the unrecognised object and therefore characterises the arrangement of the object's parts with respect to that coordinate system. Having done this the object then matches such characterisations against characterisations of objects shapes stored in a similar format in memory." The object in question is therefore described with respect to a coordinate system centred on the object itself, because even though the locations of the object's parts with respect to the viewers change as the object as a whole is moved, the location of its parts with respect to the object itself does not change. A structural description representing an object's shape in terms of the arrangement of its parts using parameters whose meaning is determined by a coordinate system centred upon that object is called the *3D model description* in Marr's theory.

Centring a coordinate system on the object to be represénted solves the problem that a single object centred description of a shape would still fail to match an input object when the object has additional parts. Such an example would be a horse with a saddle as a part. Marr and Nishihara address this *stability* problem by suggesting that information about the shape of an object is stored not in a single model with a global coordinate system but in a hierarchy of models each representing parts of different sizes and each with its own coordinate system. Each of these local coordinate systems is centred on a part of the shape represented in the model aligned with its axis of elongation symmetry or rotation.

Therefore in the case of an object such as a horse, there would be a top-level model with a coordinate system centred on the horse's torso. That coordinate system would be used to specify the locations, lengths and angles of the main parts of the horse such as the head, limbs and tail. Below this subordinate models are defined for each of these parts, one for the head, one for the front right leg etc. Each of the models would contain a coordinate system centred on that part that the model as a whole represents or on a part subordinate to that part. The coordinate system for that model would be used to specify the positions, orientations, and lengths of the subordinate parts that comprise the part in question. Thus within the head model there would be a specification of the locations and angles of the neck-axis and head-axis with respect to a coordinate system centred on the neck-axis. Each of these parts would in turn have its own model consisting of a coordinate axis centred on a part plus a characterisation of the parts subordinate to it.

Using this hierarchy of coordinate systems solves the stability problems because even though the position and orientation of a hand can change with respect to the body the position of the hand relative to the arm remains the same. Therefore the description of the shape of the arm remains constant only when the arrangement of its parts is specified in terms of angles and positions relative to the arm axis, not relative to the axis of the body. Positions lengths and angles must therefore be expressed in terms of ranges rather than precise values so as to allow for the variations in shape.

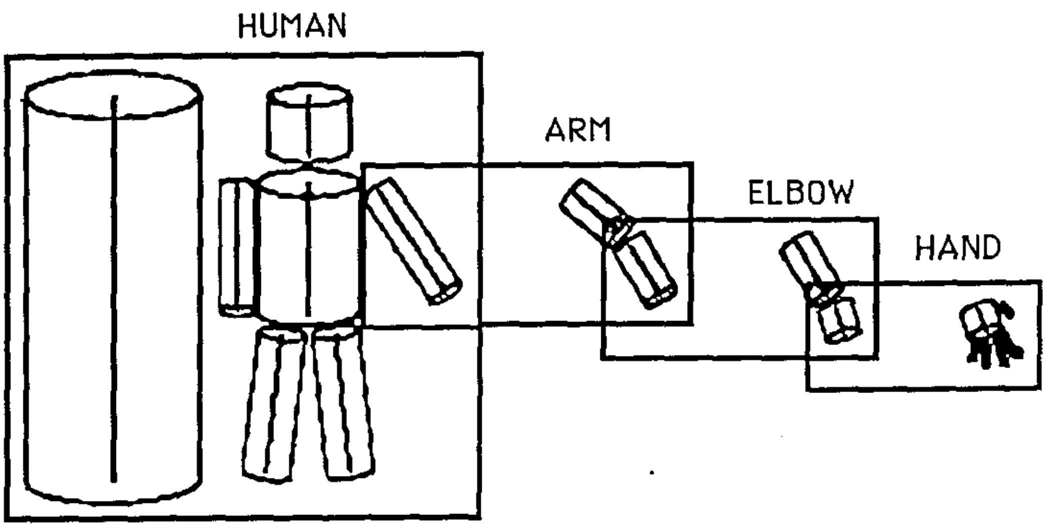
5.8.2 Organising and Arranging Shape Information in Memory

At each level of the hierarchy there is usually a well developed class of shapes which are in a particular arrangement. For example most bird shapes cluster around a particular shape. One level down two classes of shapes might be sea birds and land birds. A further level down may specify a specific bird. Information about a shape can be specified in a top-level coordinate system which is matched against models for general classes. If a shape is matched at a higher level but not at a lower level, then a subject comes to the conclusion that the object is a bird but that they do not know what type of bird. This suggests that in the analysis of pictorial information that a menu system that explicitly takes account of this fact by providing successively more detail about an object in its menu structure would be in keeping with how subjects mentally describe objects.

5.8.3 How are the Shapes of the Parts Represented?

Marr and Nishihara suggest that shapes of parts may be described in terms of generalised cylinders (Binford 1971). Just as a cone can be defined as the surface traced out when a circle is moved along a straight line perpendicular to the circle while its diameter steadily shrinks, a generalised cylinder can be defined as the surface traced out when any planar closed shape is moved along any smooth line with its size smoothly changing in any way. Therefore in order to specify a particular generalised cylinder then the shape of the axis must be specified, the two-dimensional shape of the generalised cylinders cross-section and the gradient defining how it is changes as a function of the position along the axis.

FIGURE 5.17



HUMAN REPRESENTED BY GENERALISED CYLINDERS

5.8.4 Problems with current shape recognition research

The shape primitives proposed by Marr and Nishihara - generalised cylinders centred on axes of elongation or symmetry have two advantages firstly they can easily characterise certain important classes of objects and secondly they can be easily derived from their silhouettes. However Hoffman and Richards (1984) have pointed out that many other classes of object cannot be so easily described in this scheme such as faces trees etc. They suggest that the problem of *describing parts* can be separated from the problem of *finding parts*. If parts are only found by looking for instances of certain part categories then parts that do not belong to any of these categories would never be found.

5.8.5 Orientation and Recognition of Patterns

If it were true that recognition of a pattern depended on simple features such as parallel lines and right angles then it should not matter which orientation the object in question is presented to the subject. However if recognising a pattern depended on some feature analysis with the location and orientation of the features relative to each other then rotating the pattern should slow down a subject's ability to recognise a particular shape.

Cooper and Shepard (1973) examined the amount of time subjects took to recognise letters when they were given a different orientation. They found that it took longer to recognise the letter the further away from normal the letter was rotated. This is because the further the rotation the harder it is to match the subject's knowledge of the letter to the shape in front of him/her. Therefore this suggests that orientation is important in the analysis of objects. If a subject's knowledge consists of pattern maps then recognising a pattern involves running off a series of fixations that corresponds to some pattern map in memory. However if a pattern is novel, then there is no map in memory to correspond to the pattern. Therefore in order to recognise a novel pattern the person would have to construct a map to match it.

The first fixation that the subject gets of a novel pattern would serve as the entry point into the set of maps which are in memory. The features from the first fixation would be matched with fixations in memory and the best match would serve as the basis for retrieval of a map to direct further processing. Once a pattern map has been retrieved the person could then combine this specific information with the information in the first fixation in order to determine how the rest of the pattern should look. Patterns are identified in terms of their attributes which may be lines and angles which make up the figure or they may be more complex. Figures may also be identified in terms of attributes like concavity and horizontality which are characteristics of the whole rather than simple parts.

To conclude, pattern matching is one of the least well understood aspects of pictorial analysis. However understanding how shapes are matched would help to explain how shapes are remembered and thus could be used as a technique by which to describe objects within pictures. This description of objects would be particularly useful in picture analysis. However each of the many theories attempting to explain the mechanisms behind pattern matching have explained to some degree practical examples used, yet there is no overall unifying theory which explains all patterns. Therefore the theory actually used in a practical application depends which theory is most able to account for the actual pattern observed.

As such Marr's theory in which knowing the shape, space and spatial arrangement in a scene can be distinguished from the nature of the objects and their use, is probably the most applicable in many pictorial databases because this then suggests that specifying such attributes does not interfere with the user's perceptions of the nature of the objects and their use. Thus a pictorial database which attempts to keep these two aspects of pictorial information separate would appear to take advantage of Marr's theories of pattern matching in a pictorial database environment. Additionally the use of generalised cylinders as an underlying principal by which objects can be decomposed would appear to agree with Marr's theories. Thus in cases of objects were the decomposition is not intuitively obvious then this concept of generalised cylinders might be used, since Marr has postulated that many objects are visually constructed this way.

5.9 A TYPICAL SCENARIO

It has been shown that there is a need to locate pictures from a pictorial database (chapters 2 and 3). A subject searching for visual information from a pictorial database system, will normally make use of mental techniques such as imagery and memory in order to bring into conscious attention the attributes of the objects and/or the spatial relationships between objects within a desired picture. If the subject has not seen the picture before then only general memory and imagination can be used to construct database searches. However if the picture in question has been viewed before then remembrance of the properties of the specific picture will play an important part in locating it. Here what the subject remembers about the objects and attributes of an image can be used to recall specific pictures. When viewing pictures located by a query, a subject will need to be able to detect the one that matches the recalled picture. Clearly recognition has an important function in this kind of search. This chapter has suggested that when a picture conforms to a real-world scene then the visual and spatial components of a picture distinguish one picture from another and that the information searcher can search data in terms of these properties. In general we suggest that users may want to conduct such searches with respect to any pictorial database.

In the event that an image has been viewed in the past or the user has knowledge of the objects and relationships, then the user might be able to construct such an example from memory and be able to physically draw a crude representation of the image remembered. Psychological evidence suggests that the extent to which it is possible for the user to search a pictorial database in terms of spatial and visual attributes of the data in cases where queries cannot be constructed from examples, will depend on the human potential for storing, recalling, imaging and recognising visual and spatial information.

In identifying visual information from a pictorial database three of the most important physical aspects are psychological imaging, visual memory and recognition for the objects and spatial relationships contained within the pictorial information. Practically we suggest that a visual querying system entails locating the required objects either by selection from a variety of visual representations, selection from a textual menu or a combination of both of these techniques. In selecting the required object from a choice of visual representations we suggest that a pictorial menu can be used in which graphics convey information as to the nature of the object representations required. Alternatively such objects could be described using textual descriptions, and again a menu of these textual descriptions could be selected from. Having selected an initial pictorial or textual representation, further details as to the nature of the object being searched for can then be specified, either using more detailed pictorial representations or additional textual descriptions. Here then it is suggested that a hierarchical structure may be used in the query formulation by arranging the information to be searched for in increasing levels of detail.

This hierarchical structure utilises the user's ability to spatially navigating through a database management system specifying details about the object in increasingly greater detail as they hierarchically progresses through the system. In the case of pictorial representations, specifying such data is facilitated by the distinctive appearance of the data and the route by which it is accessed. Using this technique users can be progressively more detailed about the nature of the objects they require. Such a hierarchical selection technique also allows users to browse the database by enabling each of the actions in the progressively more detailed selection to be easily reversible. Ideally some form of pointing device such as a mouse should be used providing a simple and natural means of moving through the database enabling users to explore the entire database with one easy to use control. This technique acts as an alternative to conventional databases where users must pose successive queries in order to obtain information.

Having identified the object of interest, visual attributes such as the spatial location of the objects and their size within a picture can then be detailed. It is suggested that these attributes can be specified either by visually depicting the chosen objects (for example by specifying the spatial location, size and orientation of the objects) or alternatively by textually entering the numeric values of the attributes of the objects in question directly (for example in the case of the physical object attribute of orientation, then the value 30 degrees can be textually entered if it is wished to find an object at an orientation of 30 degrees to the horizontal).

Having specified both the object of interest and the attributes associated with such objects, then this information can be compared with data directly abstracted from pictures contained within the pictorial database in order to identify matching pictures to the user's query. Such a system would allow users to indicate the information about a picture that is known, either for the purpose of recalling a picture that they have seen before, or alternatively as a means by which to identify a picture that matches user specifications. Therefore a combination of imagery, memory and recognition is used in identifying the objects and attributes of interest. This depiction of pictorial information as a means of selecting pictures is further discussed in the next chapter.

5.10 CONCLUSIONS

5.10.1 Conclusions about Spatial Relationships Within Pictures

Many tasks require the use of knowledge that either pertains to or is derived from objects in either two-dimensional or three-dimensional physical space. Such knowledge is referred to as spatial knowledge. This spatial knowledge can be broadly subdivided into analogical and propositional representations. Analogical representations are useful for encoding spatial configurations of objects but less useful for encoding semantic or logical knowledge. Propositional representations store spatial knowledge in the form of textual descriptions however these run into difficulties the more numerous the spatial relationships become. The proposal in this thesis is that spatial attributes can be queried and examined analogically rather than propositionally assisting users in querying pictorial databases by using their visual memory of object positions and spatial relationships with other objects.

The encoding of spatial information is influenced by the intention to learn. If the user comprehends that a system relies in part upon their ability to recall spatial information and that the recall of this spatial information more effectively enables them to locate information that they require, then surely the intention to learn spatial information associated with a picture will be enhanced. Such encoding of spatial location information within the brain is influenced by competing task demands. This suggests that where pictures are extremely complex in their appearance then the subject's memory for spatial arrangements of such objects is significantly decreased. This further suggests that the more complex a picture is, the less important spatial relationships within such a picture become and the more emphasis should be allowed to be placed on other visual attributes within the picture such as size, orientation and colour, thereby enabling the user to locate information based in part on these other visual attributes rather than purely on a spatial method of identifying pictorial information.

The encoding of spatial information and the subsequent querying of this spatial location information is influenced by practise.This suggests that the more a subject uses a pictorial database management system in order to locate pictorial information by specifying the spatial location and spatial relationships within the pictures, then the greater will be their effectiveness in locating the correct information from it.

Psychological results imply that neither absolute nor relative locational information are automatically encoded. However it has been suggested that the coding of spatial information consists of a continuum linking automatic and non-automatic processes and that there is a greater than incidental memory of spatial information when viewing a picture. It is this fact that there is a greater than incidental memory of the spatial location of objects that suggests that visual spatial techniques of search may be used to locate pictures from a pictorial database. Since this information is stored analogically rather than propositionally within the brain, a system which allows users to query pictorial databases depictively rather than descriptively should allow them to more closely match the form of their query to the information stored within their memory.

Conclusions about the arrangement of spatial information are that our knowledge of spatial relationships is not knowledge drawn directly from memory; rather it is the consequence of our ability to manipulate stored information to produce appropriate and novel responses. Therefore in a pictorial database containing visually different objects, it is their position rather than spatial relationships between them that are remembered. Thus it may be best to base searches on spatial information such as the location of an object within a picture (top-left, bottom-right etc) rather than requiring the user to specify how the objects are spatially oriented to each other (such as an object to the left of another object, or below another object).

Given a specific object upon which to base the location of other objects, subjects recall other objects in a top-to-bottom fashion, rather than specifying the nearest objects first. This suggests that within a pictorial database in which pictures are identified based upon their spatial position relative to a specified object in a picture, users should be asked for other objects in a broadly top-to-bottom arrangement. Therefore in the identification of a picture users should be prompted for the spatial position of the object nearest to the top of the picture first and the object nearest to the bottom of the picture last.

Since the ordered tree produced from subjects recall shows that objects that might be considered to belong to the same semantic class such as book, envelope and ruler were not grouped together in memory but were grouped upon their spatial proximity, it is suggested that this strengthens the proposal that objects within a picture can be searched for by specifying their spatial positions or spatial relationships to other objects, since this appears to be how information is stored in memory. If the ordered tree produced from subjects had shown that objects were grouped semantically this would have suggested that an important factor in the depiction of information within a picture would be those queries in which objects of the same semantic class were depicted together. The findings from the ordered tree also argue that spatial memory has a hierarchical component, even when physical and perceptual boundaries are absent and therefore that complex pictorial information is spatially subdivided into 'groups' of adjacent objects, these groups being further subdivided into smaller groups.

However it has been found that subjects tend to group identical objects into higher-order units and store these units rather than individual objects. Therefore similar objects are grouped into patterns. These patterns usually occur between neighbouring objects. This implies that where objects within a picture are identical then it may be best to allow users to search pictures based upon how the objects are spatially oriented to each other rather than specifying their location within a picture. Therefore within a pictorial database in which identical objects are present, queries about the spatial arrangements of the objects should be in the form of the relative positioning among the objects rather than their locations within the picture.

5.10.2 Conclusions about Pictures

The ability to understand pictures is learned at a early age. This ability enables people to comprehend the contents of pictures even though they may never have seen the objects contained within them themselves. People learn to make assumptions about pictures based upon this experience since in most cases these assumptions are correct, only in examples such as optical illusions are they wrong. Therefore it is suggested that these assumptions about the content and nature of pictures can be used to locate information from a pictorial database.

The speed at which pictures can be seen and interpreted is faster than with text. This is due in part to the fact that pictures have many striking visual characteristics such as brightness, colour, contrast and size. People tend to fixate on areas that are highly informative within a picture and one of the factors that determine the scanning pattern followed by a subject's eye is the subject's purpose in looking at the picture in the first place. Therefore it is suggested that these highly informative areas are important attributes of a picture and as such can be used as a means of identification in a conceptual system which allows the user to visually depict an object query for the purpose of finding a picture.

It has been reasoned that there are at least three kinds of memory-relevant information a subject could extract from a picture which might be influenced by the nature and organisation of that picture. These are the theme of the picture based upon an inventory of the objects within it, the objects and their visual attributes and the actual location of the objects as well as the spatial relationships between them.

Within a picture subjects tend to visually process pictures for theme, rather than the exact visual details contained therein, in effect there is partial recognition. The human brain tries to produce patterns from random selection of pictorial data wherever possible. It appears that pictorial memory needs the ability to use organisation to construct meaningful representations of integrated visual ideas in order to achieve maximum effect. Visually organising pictures helps users to remember the spatial information in the picture and thus the actual objects present. There is also a tendency to make unorganised pictures conform to organised scenes which suggests a loss of memory for the specific pictures and the use of a real-world organisations to guide construction at the time of recall.

Many objects within a pictorial database can only be readily identified by using pictorial representations. For example faces are extremely difficult to textually describe but are usually immediately recognisable when seen. This suggests that pictorial representations are the most useful way by which such information can be located. Additionally where objects within a picture are visually distinctive, then they are easier to remember especially if such pictures contain striking or unusual objects. For example if a whale was present in an everyday street scene then this would be sufficiently unusual to be visually memorable.

Five important types of constraint imposed by pictorial organisation have been found in which the nature of the picture is sufficiently unusual to be memorable, these are support, interposition, probability, position and size. Each of these constraints when broken makes a picture more memorable to the user. However when objects within a picture become totally disorganised then subjects are unable to base their recall of such pictures upon real-world schema. Therefore for example a scene which contains a person flying through the air would be memorable but a scene in which all objects were arranged completely randomly would be less memorable since the user would not be able to make assumptions about the spatial positioning of the objects present.

People often use such locational information to help organise objects for recall and it appears that a large quantity of locational information is coded into long term memory in the sense that active processing of this information is not required. People's positional memory about a scene appears to persist long after the actual knowledge of the objects within it becomes hazy. Even when subjects are instructed to remember particular attributes of objects within a picture, such as colour or shape the memory for these features is not as efficient as spatial memory. Therefore this suggests that positional information is important and should be encoded in some way into picture analysis systems enabling a subject to more readily identify pictures based upon the spatial arrangement of the objects present within them.

The spatial concept of locating information is therefore utilised by arranging the data in increasing levels of detail enabling users the ability to control their steps in order to achieve their goal. Applications of such systems rely upon the user's spatial ability in navigating around such a database. Locating information with the aid of hierarchical menu structure is best for those queries in which the specific formulation of a query is difficult. In general spatial database management systems deal with large heterogeneous collections of spatial and non-spatial data which require flexible methods for interactive inquiries. Actions within such structures where possible should be easily reversible. Finding information visually from such systems is facilitated by the distinctive appearance of the data. Having identified a specific picture the technique of pattern recognition enables subjects to recognise objects. Therefore there must be some kind of 'symbol' stored within our brain which can be used to stand for a visual feature or object and constitute its explicit description. It is suggested therefore that this symbol must be sufficiently generalised to match many objects which users would consider as constituting the fundamental aspect of the object in question. Therefore this suggests that the initial choice of representations from a pictorial menu in the identification of objects from a database should contain representations which are sufficiently generalised so as to match the user's perception as to what constitutes the object required. For example the initial pictorial representation of a chair would therefore be as simple as possible with four legs, a seat and possibly a back, rather than a more elaborate construction such as a throne which contains many elaborations which are not strictly necessary in order to constitute a chair. A more detailed example of a chair which more closely matched the user's perception of a specific example could then be identified using a combination of hierarchical menu systems and the specific attributes of the chair in question.

At each level of the hierarchy within pattern recognition there is usually a well developed class of shapes which are in a particular arrangement. Therefore this suggests that practically a system which enables users to be increasingly more specific about an object of their choice would therefore match their pattern recognition ability. In appendix 1 both representations of objects (icons) and direct manipulations of such representations are examined. Examples of their applicability to pictorial database systems is also examined.

To conclude since there is a requirement for users to locate pictorial information (see chapter 2 and 3) and that both spatial and visual attributes of the pictorial information are remembered, then it is suggested that a pictorial database management system can be constructed which attempts to take account of the user's visual memory of both the appearance of the objects and the visual attributes of the objects within the picture required. Such a system should allow the user to both peruse and manipulate information in a lifelike manner by using both graphical interfaces and simple navigational controls (see chapter 4). Within such systems it is suggested that queries can then be depicted in a manner that is similar to how the images are mentally stored within the brain. An additional feature of such a system is that the ability to formulate imprecise spatial queries should be possible since 'ranges' of values allow the user to locate pictorial information based upon only vague memory of the spatial location of objects within a picture.

CHAPTER 6

PROPOSALS FOR THE ANALYSIS OF A PICTORIAL DATABASE BASED IN PART UPON A SPATIALLY DEPICTED SEARCH

6.7 POSSIBLE APPLICATION AREAS FOR SPATIAL SEARCH TECHNIQUES

6.7.1 Military Applications

6.7.2 Cells

6.7.3 Paintings

6.8 CONCLUSIONS

6.1 INTRODUCTION

In this thesis pictorial information stored on both non-computerised and computerised media have been analysed, in addition to spatial techniques of locating pictures within a conceptual three-dimensional space. A psychological analysis of people's ability to both recall and remember spatial information within pictures and their ability to interpret pictorial information has also been studied. This chapter uses the conclusions drawn from all these studies to propose a database management system in which the user can locate pictures from a pictorial database by visually depicting both the pictorial representations and the spatial locations of the objects within them. The first stage in this proposal is to use the conclusions about visualisations of pictorial information to verify that such a proposal is reasonable. Having used psychological evidence to verify such a proposal, the concepts and requirements of the proposed system are then examined.

6.2 THE VISUALISATION OF PICTURES

Human visual mechanisms have developed to understand the three-dimensional real world in which we live. That world consists largely of objects, surfaces and their coverings. For example when we glance at a scene, we gain an immediate appreciation of its three-dimensional surface structure, what the surface is covered with and even the condition or state of the surface covering. This ability to shift attention or reconfigure mentally from appreciating one physical property of the scene to another, is critical in allowing us to understand scene properties (see chapter 5).

This explains why, when we are faced with the task of finding appropriate data variables for the visual representations of an object or scene, representing objects or scenes by their physical properties can be so successful. Visual mechanisms which are known to be efficient can be exploited in understanding the structure and components of an object or scene (see chapter 5), furthermore Clarisse and Chang (1986) have suggested that visual representations display information in a form which resembles the way that people think. This provides initial guidance to

developing a methodology for choosing appropriate data representations of such information. Such techniques can be most powerful if they present data in a form which exploits natural and well-developed visual abilities for the appreciation of attributes such as spatial structure. Underlying the concept of visualisation of pictorial information is the idea that an observer can build a mental model, the visual attributes of which represent data attributes in a definable manner.
We suggest that these visualisation techniques can also be applied to two-dimensional abstractions (ie pictures) of objects or scenes, this being a type of information that the human visual system is experienced and adept at interpreting (Robertson 1988). Pictures also exploit the ability of the human visual system to recognise spatial structures and patterns. We also suggest that the application of visualisation techniques is particularly useful in locating pictures from large pictorial databases.

Reasons why a searcher might wish to locate such pictures include:
Searches for pictures which have been previously selected and are known to be in the pictorial database or alternatively search for pictures which are not known to be in the pictorial database. In the second case there are several possible reasons for searching, either to confirm that the pictures exist, to confirm the non-existence of pictures or to ensure that all items meeting a certain criterion are found (such as how many pictures have a requested object in them). In order to locate such pictures from non-computerised sources, systems have been developed which attempt to base the search upon the objects within the pictures (chapter 2). To date however there are hardly any proposals that allow users to do such content-search of computerised image data (Meyer-Wegener, Lum and Wu 1989).

In the past the most common method of communicating with computers has been through the medium of typed words at a keyboard, unfortunately this has many problems foremost of which is the fact that people often think in terms of pictures rather than words (Lansdale 1988). Users must then attempt to translate these pictorial thoughts and ideas into textual statements for later input into a computer. Unfortunately under these circumstances many of the original visual thoughts and ideas have to be sacrificed in order to express the information in a form that the computer can readily understand.This in turn leads to restrictions in effective communication between the computer and the user. It is therefore suggested that locating pictures based upon visual representations of the objects within a picture more closely matches the mental representation of a picture than any form of purely textual query and that a computerised database management system can be built which uses these visual representations of objects as a search criterion by which to locate pictures.

An important step in choosing effective data attributes from our visualisations of pictorial information for the incorporation into a database management system, is to look for guidance from our interpretations of pictures. Two of the most important techniques in the analysis of pictures are our ability to interpret three-dimensional information from pictures and our ability to analyse pictures based upon their organisation. A critical study of both these visual techniques follows.

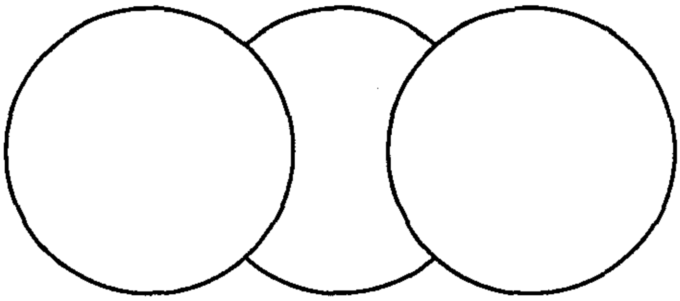
6.2.1 Interpretation of Three-Dimensional Information

In real life objects observed by the eye are of three dimensions. Solid object surfaces extend away in the distance from the eye. The faces of these objects can also be inclined at different angles to the eye such that only a proportion of the object in question is visible.

Pictures are two-dimensional representations of three-dimensional objects and as such can not display the quantity of information that a three-dimensional scene displays to the eye. However this has to some extent been overcome by the pictorial cues of perspective and relative size within pictures which give the illusion of depth. Thus pictures may differ in a number of conceptual dimensions and still adequately fulfil their symbolic purpose.

One of the most important aspects of a picture is its ability to display only part of an object and yet still allow the subject's perception of the object to be perceived as a whole rather than a part. For example figure 6.1 appears to consist of three complete circles rather than two circles and sections chopped out of the third. The ability of people to fill in missing information absent from a picture, is however one of the hardest mechanisms to take account of in computerised image processing systems. Unlike people, who instinctively rely on their experience of the nature of objects, computers are normally unable to extrapolate pictorial information which is not present.

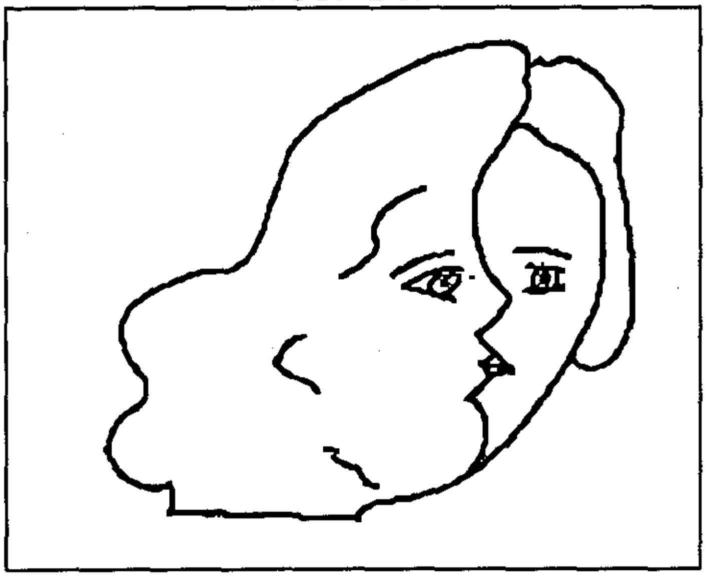
FIGURE 6.1



PERCEPTION OF THREE CIRCLES

In the case of the analysis of two-dimensional projections of three-dimensional scenes, it is suggested that objects should be classified as they are seen, rather than as they are perceived, since this avoids the problem of attempting to take account of hidden visual information. Thus under these circumstances only information that is displayed within the resulting picture should be referenced, not information that is partially or completely obscured by objects within the picture conceptually closer to the viewer. For example in the picture shown below one of the two faces is partly obscured by the the other. Assumptions made about the partially obscured face (such as shape) may differ between subjects observing the picture. In order to classify this information, a different projection should be taken in which the previously hidden information can be clearly seen. People using pictorial databases should always be aware that any picture only represents one view of a scene, and obtaining more information about the nature of the scene will require several views to be taken from different angles.

FIGURE 6.2



TWO FACES ONE PARTLY OBSCURED BY THE OTHER

6.2.2 Analysis of Pictures Based Upon their Organisation

It has been shown that pictures can be analysed in three basic ways, each of which are influenced by the organisation of the picture (Mandler and Johnson 1976).

1) Firstly by examining the overall picture.

2) Secondly by examining the objects and their visual attributes in a picture.

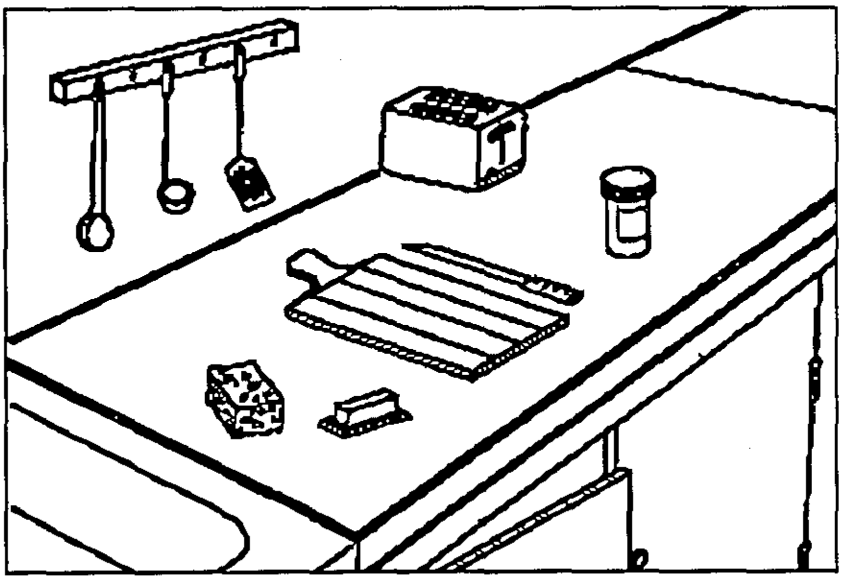
3) Thirdly by examining the spatial relationships between the objects in a picture.

6.2.2.1 Overall Picture Description

Pictures can be categorised by a textual description of the scene present within the picture (Pezdec 1978). For example in the picture below a kitchen is shown in which there is a number of objects spatially arranged within it. Here the nature of the objects and their spatial arrangements help a subject to decide what the picture actually represents. The classification is of the grouping of objects and relationships within the picture rather than of any one specific object present. Although there may be several prominent objects within the picture, it is the general impression of the picture that is important for an overall textual description. This description can not easily be extracted by an image processing system (a computer would have difficulty in determining that the scene below represented a kitchen!).

If the theme of a scene (the word kitchen in this case) is used to locate pictures from a pictorial database, then the naming of such scenes must therefore be manually specified by a database management designer enabling the user to later select the scene based upon such a name. This description however is subjective and there is no rigorous scheme by which the naming of such scenes can easily take place (see chapter 2). Under such circumstances the scene's description is based upon what the database management designer suggests that the scene represents rather than any systematically rigorous description. This implies that such a description should not be used as a search mechanism to locate pictures since there is a possibility of a mismatch between the database designer's perception of the theme of the scene and the user's perception. The scene can however be decomposed into its constituent objects. For example the picture of a kitchen in figure 6.3 might be decomposed into consisting of the objects: toaster, breadboard, knife, spoon, etc. An analysis of the relationships between such objects follows in the next section.

FIGURE 6.3



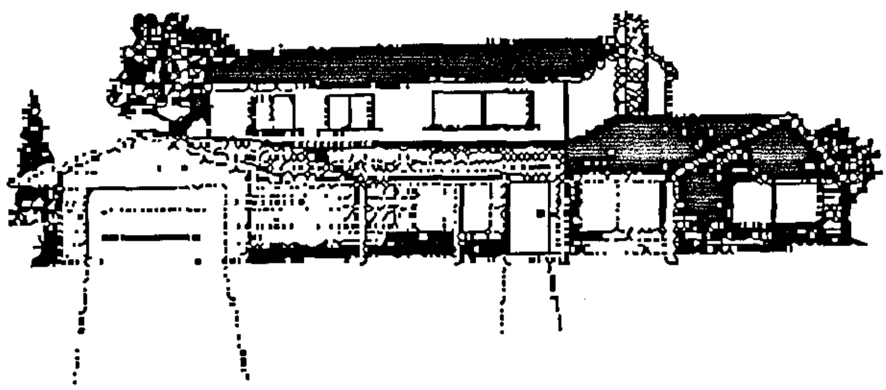
PICTURE OF A KITCHEN

6.2.2.2 <u>Object Classification</u>

Mandler and Johnson's (1976) second classification of a scene is in terms of the objects and their visual attributes. However the complexity of many object's visual appearance, means that they can often be decomposed into simpler ones, both for classification purposes and to allow the user to 'home in' on the object (or part of an object) of their choice (IBM,1986). This object decomposition is the process of defining for each object the objects it consists of, and the relationships between its component objects contained within the lower hierarchy.

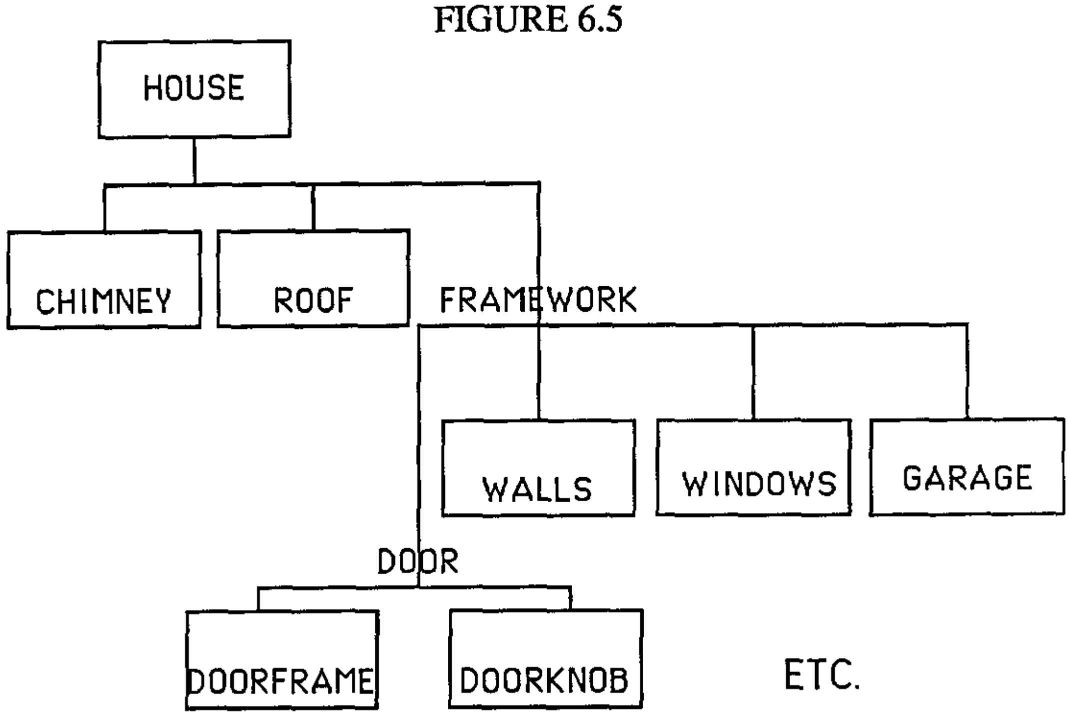
An object can therefore be thought of as an image itself and can be decomposed into other objects which can be considered to be at a lower hierarchical level. Objects at a lower level can be further decomposed into other objects and so on until the simplest object within the hierarchy is specified (Economopoulos and Lochovsky 1983). It is suggested that identification of pictures can be based (in part at least) upon these component objects within a scene. The example in figure 6.4 shows a picture of a house and its hierarchical decomposition into component objects. This house consists of walls, windows, a chimney, roof, door and garage which could be considered to be connected hierarchically as shown in figure 6.5.

FIGURE 6.4



PICTURE OF A HOUSE

FIGURE 6.5



HIERARCHICAL DECOMPOSITION OF THE HOUSE INTO ITS CONSTITUENT OBJECTS

However in pictorial database management systems in which pictures are classified solely in terms of the names of objects within them, then in cases where similar objects are present within many pictures (such as doors), then a search based upon only the names of these objects will produce many pictures for users to search through in order to find the picture that they require.

It is therefore suggested that the most suitable description of pictures is one that includes both a classification of the objects within the pictures along with physical attributes of the objects such as the spatial information associated with them. This can be seen as a combination of the ways that Mandler and Johnson have found of analysing a picture, since other visual attributes in addition to spatial information are included in the classification. This thesis is however primarily concerned with the spatial information associated with objects within a picture and it is this attribute that will be examined in greater detail. The memory for the spatial information associated with an object's position is under many circumstances at least as memorable as the objects themselves (Mandler and Johnson 1976). The term 'spatial' information applies to any data concerning phenomenon distributed in two-, three-, or N-dimensions (Peuquet 1984). Many existing pictorial database management systems use a limited set of single built-in textually based descriptions in order to describe spatial relationships between objects, however the ability to handle multiple spatial relationships in single textual queries is very restricted. This means that complex spatial queries must be constructed by the user through a sequence of simpler textual statements (Pizano, Klinger and Cardenas 89). For example the textual classification of spatial relationships between objects within pictures using textual descriptors such as LEFT-OF, RIGHT-OF, ABOVE and BELOW may produce numerous pictures for users to view from which a manual search may be required for the users to find the picture that they require.

More advanced pictorial database systems define spatial relationships in terms of simple geometrical models, however these are unfortunately only approximations of how a human would interpret them in most given situations (Peuquet 1988). For example to textually describe the spatial relationship between three objects in a picture as triangular would only be a crude approximation of such a relationship since there are many shapes and sizes of triangles. The ability to deal with complex ad-hoc spatial queries in order to identify specific pictures from a database is important in picture identification since such an ability should allow users to more accurately identify a picture that they require by specifying such a query. This is especially true where there are a large number of pictures present within a pictorial database in which numerous spatial relationships exist.

Practically, spatial database management systems have shown that the concept of spatiality within three-dimensional space can be used to locate information from within a pictorial database (see chapter 4). Within these systems graphics are normally used to convey information (often in the form of icons) and users can visually select information without a detailed knowledge of the database structure. Simple concepts such as panning and zooming are also included which allow the user to navigate around the system. Using this spatial concept of identifying information, it is suggested that visual information such as an object's representation and position can be depicted on the user-interface, after which this depicted information can be matched to information within a pictorial database. Both subjective and experimental evidence suggests that humans prefer to interact with a system through a medium which exploits a two or three-dimensional dialogue (eg Schneiderman 1982, Bolt 1984).

In general when observing a two-dimensional picture the spatial attribute, consisting of the spatial positions of objects within the picture and the spatial relationships between these objects has been shown to be one of the attributes that is remembered without conscious attention having been paid to it (chapter 5). A locationally based-query can in general be described as: one that finds the given objects that satisfies the required spatial constraints. These spatial constraints can ultimately include a locational window for the position of an object and any combination of spatial relationships with other objects. Stated in this way it can be seen that spatial relationships can be used directly by a pictorial database as a search constraint on locationally organised visual information (Peuquet 1988). It is suggested that matching two-dimensional spatially arranged representations of objects on the user interface to pictorial information within a pictorial database, would (in part at least), overcome the problems of textually querying such a database since using this technique users should be able to locate specific pictures, based upon their visual memory of either the positions of objects or the spatial arrangements between objects in a picture.

Visually depicting spatial information for objects which differ in their appearance is however best achieved by depicting the positions of the objects within the picture rather than their spatial arrangements with other objects, since for objects which differ in appearance it has been shown that subjects build relationships between objects from their knowledge of the location of the objects within the picture (see chapter 5). Further evidence for the concept of visually specifying spatial queries comes from the ordered tree produced from subjects recall, showing that objects are grouped in memory based upon their spatial positions rather than their semantic groupings (chapter 5). It therefore seems reasonable to attempt to construct a database management system in which visual attributes can be depicted, one of these attributes being the absolute spatial position of objects within pictures.

Such a system would conceptually consist of the following:
The user first selects or sketches pictorial representations and the values of object's visual attributes that the system understands and enters them on to the user-interface.This enables the database management system to carry out the appropriate matching procedure and select a picture (or pictures) from the pictorial database containing the objects which match to the spatially positioned pictorial representations and values of visual attributes entered by the user. These matching pictures from the pictorial database are then displayed. However due to the practical problems of sketching pictorial representations (Herot 1986), it is suggested that representations should be primarily chosen by selection from menus rather than by sketching detailed visual queries.This combination of selection of object representations and the visual specification of physical values of attributes such as spatial location, should provide the necessary information for the identification of objects and pictures present within a pictorial database. In order to select object representations and place such representations on a user-interface two important facilities are required.

1) The ability to select object representations from a menu and

2) The ability to spatially place object representations on the user-interface.

A study of these two proposed facilities and their concepts now follows.

6.3 SELECTION OF OBJECTS REPRESENTATIONS FROM A MENU

6.3.1 Physically Selecting Iconic Representations of Objects

Here iconic representations are chosen which match to objects within pictures. It has been shown that icons are most effective as miniature representations of the physical objects to which they refer (Rogers 1986). Advantages of finding objects by selecting iconic representations of them include the fact that screens of icons can be easily visually examined (Hirakawa, Monden, Yoshimoto, Tanaka and Ichikawa 1986) and that an icon interface may substantially obviate the use of other media (such as commands) and provide a more usable dialogue because of its capacity to to carry much greater descriptive information using the same, or less physical display space (Gittins 1986). A detailed study of icon classification is examined in appendix 1.

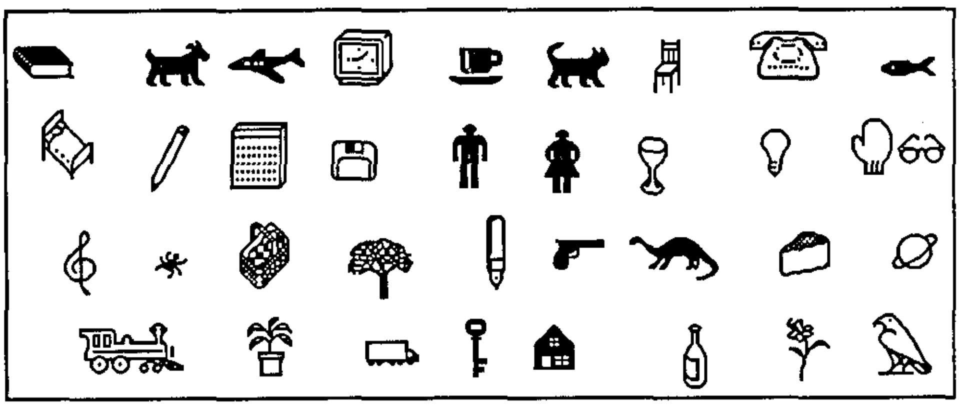
6.3.2 Generalised and Specific Icons

The well-established method of icon design capitalises on their information-carrying capacity by unifying individual icons into a collective metaphor (Gittins 1986). Thus the icons on an interface can have a form which corresponds to actual objects in the real world with which users are familiar. Icons can serve many different functions, however it is suggested that finding objects within pictures based upon iconic selection has two basic forms. Either the icon can be specific to a chosen object, or an icon can be regarded as representing a 'class' of objects from which a more detailed selection can be made.
In the case of the specific icon, selection should lead to the display of a specific matching object from within a pictorial database. For example an iconic

representation of Abraham Lincoln would identify all pictures within the database containing this American president rather than all pictures containing men (here men can be considered as a class).

An alternative to the specific icon is an icon which represents a class. Icons which represent a class rather than a specific example, may require users to later make a more detailed specification of the physical representation of the object that they require. For example the iconic representation of a car could be selected after which its attributes (such as the car model) are specified in more detail. This selection of a 'class' icon may thus either lead the user to a more detailed iconic selection within a different iconic menu, or may require some form of detailed textual entry in order to find a specific object present within a picture. Icons which represent a 'class' of objects are more commonly called *generalised icons* (Chang 1987). Therefore a generalised icon can be defined as representing a class of objects rather than a specific example. A generalised iconic menu is shown in figure 6.6.

FIGURE 6.6

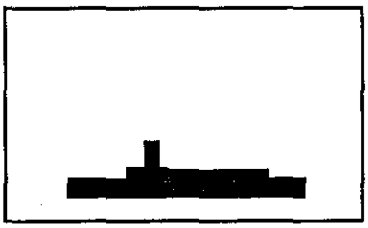


GENERALISED ICONIC MENU

In order to use a generalised icon to retrieve a specific object from a pictorial database, such as a picture containing a Siamese cat, then the initial icon shown within the iconic menu would be that of a 'generalised' cat representing all cats rather than an iconic representation of a specific example.To find more specific examples of cats rather than all pictures containing cats, further details of attributes of the cat would then need to be specified such as size, shape, colour, length, etc.

The choice for a database management designer whether to chose generalised icons or specific icons as pictorial representations of objects within the pictorial database will depend in part on the nature of database used. Pictorial information within a database may exist in the form of many different kinds of object or a more restricted domain of similar objects where the physical grouping of different icons on the display can be used to indicate some attribute which is shared (Gittins 1986). Where a restricted domain of pictures exists, then more detailed iconic representations of objects may be required within an iconic menu. For example for two pictorial databases, one containing pictures of forms of transport and the other only pictures of ships, then in the first case a generalised icon of a ship would be sufficiently detailed to identify pictures of ships such as that shown in figure 6.7.

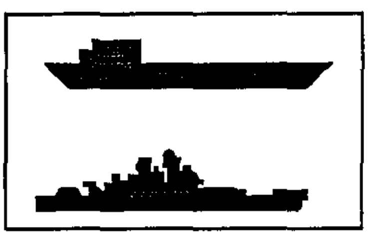
FIGURE 6.7



SIMPLE GENERALISED ICON OF A SHIP

whereas in the second case specific iconic representations of ship types would be required to more narrowly define the form of the ships required, these are shown in figure 6.8.

FIGURE 6.8



SPECIFIC ICONIC REPRESENTATIONS OF SHIPS

6.3.3 Topological and Real World Pictorial Information

It is suggested that both generalised and specific icons can, in general, be matched to two sorts of pictorial information, either:

1) Matching the icons to objects within a real world image (such as a photograph). Here the database management designer textually encodes the information pictures into a knowledge base and the pictorial database management system compares the user's iconic query with this knowledge base thereby locating matching information.

or secondly

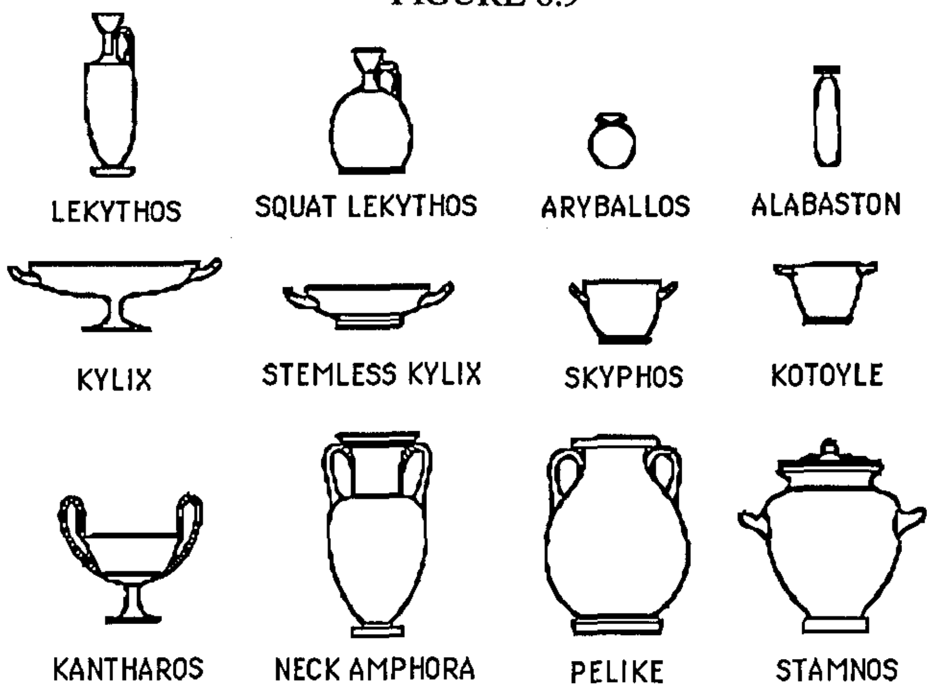
2) Matching the icons to a topological representation of a picture (such as a map). Here the match between the icon and the topological representation is considerably easier because the topological representation is already an abstraction of data from a real world scene. However this means that much information may have been lost in the initial transformation from data to topological representation, and yet further information lost in the encoding of topological information into the database management system for later comparison with the iconic representation. Therefore under these circumstances the user is more restricted in the information that can be abstracted from such a system.

6.3.4 Picture Descriptors

In order to compare both users object queries (in the form of icons) and depicted attribute queries (such as colour, size, etc) with objects in a picture then either image analysis techniques (enabling direct picture matching) or numerical or textual picture descriptors (enabling comparisons of descriptions between objects and their representations) are required. However since image matching techniques require sophisticated analysis of both the appearance of the icon and possible matching objects (Bruce and Green 1987) then it is suggested that textual or numeric descriptors should be used for the comparison of object queries with information in a picture. Choosing such descriptors is however difficult because of the enormous variety of object representations and attributes that objects can have which identify them. These attributes include general ones such as shapes, sizes, colours, orientations as well as other attributes specific to the objects being searched for.

If the objects within the pictorial database all belong to the same general class then the selection of which object attributes to chose for unique object identification is considerably easier. For example Hollerbach's analysis of vases categorises all vases to be a variation of a generalised cylinder consisting of a vertical axis along which is projected a two-dimensional shape (Winston 1979). These vases are all considered to be variations of simple two-dimensional generalised cylinders, and it is this property which enables them to be either visually or textually differentiated. Therefore when all objects within a pictorial database are of the same class, analysis and identification can be carried out by examining a characteristic attribute of the objects in question. In the case of the vases in figure 6.9 below they can all be seen to consist of a generalised cylinder which is rotated about the vertical axis and this is the means by which Hollerbach classifies them.

FIGURE 6.9



VARIETY OF VASES CLASSIFIED BY THEIR GENERALISED CYLINDERS

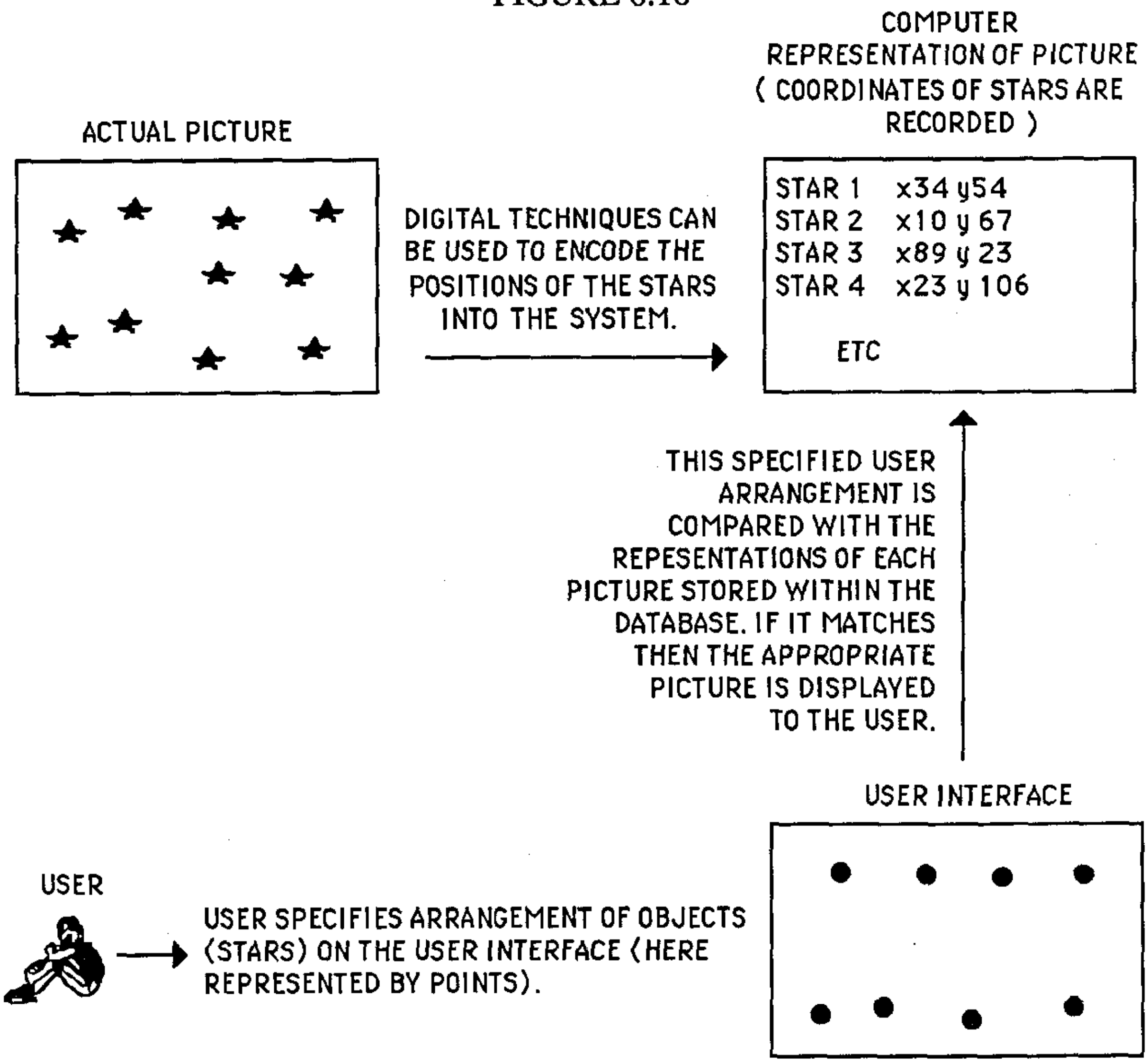
However where images are more varied containing a variety of objects then evidence from psychological studies needs to be applied in the textual encoding of images for the later comparison of this abstracted information with the user's query.

6.3.5 Encoding of Images

There are two basic alternatives by which encoding of images can be achieved. If the pictures contain simple objects, then the encoding of the attributes such as the positions of the objects can be executed automatically using the techniques of digital image analysis. Digital image processing techniques although still in their infancy are often able to detect and identify individual objects within images provided that the domain remains restricted (all pictures containing similar objects) and the objects themselves are relatively simple (Haralick and Simon 1980). One pictorial database which might be analysed using this technique is a pictorial database of stars.This system is represented schematically in figure 6.10. Here, since stars can be represented as point objects and there is a marked contrast between stars and their background in a picture (usually white stars on black backgrounds), then digital analysis techniques can be used to encode the positions of stars in a picture into a database management system. Such automated image understanding systems increasingly rely upon expert knowledge to help analyse objects and control the analysis process.Thus where the pictorial database contains information which can easy analysed then it is suggested that digital analysis techniques can be used to encode attributes such as the object's positions within a picture for later comparison with the user query.

In more complex pictures (such as real world scenes) containing objects which differ in shape, size, orientation and colour, then the values of the attributes of objects usually need to be manually encoded. This is because digital analysis techniques are not yet sufficiently advanced enough to be able to identify complex objects from images. The problems of object encoding will largely remain until digital analysis techniques are sufficiently advanced enough to be able to extract information from a greater variety of picture sources (Herot 1986).

FIGURE 6.10



DIGITAL TECHNIQUES USED TO ENCODE OBJECTS POSITIONS

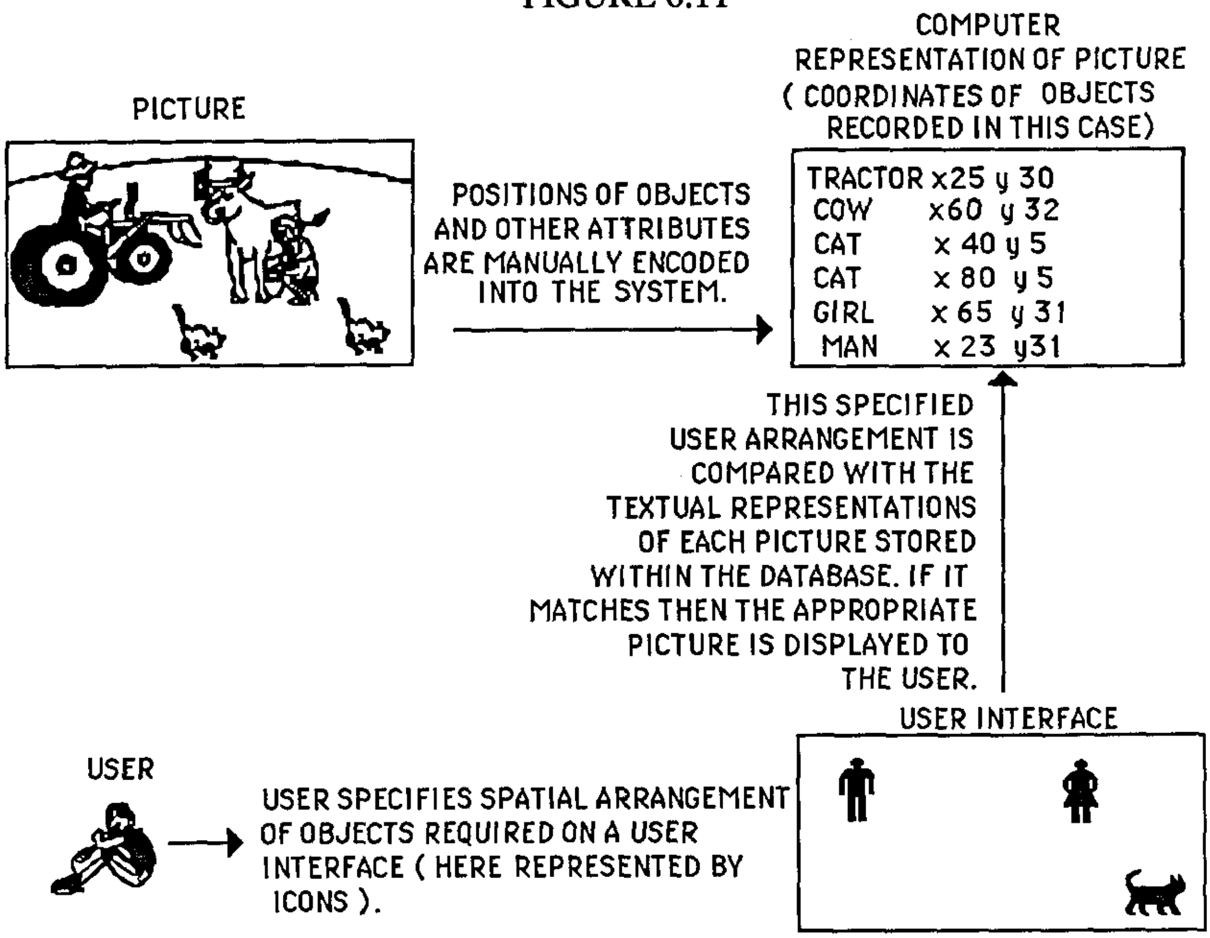
There are however significant problems with manual techniques of encoding object descriptions and corresponding physical attribute values. The major problem here is which objects and which attributes to encode into the database management system enabling users to locate pictures based upon this encoded information. For pictures in which a large number of objects are present, the database designer may need to make a subjective choice as to which objects and what attributes of those objects are actually encoded into the system. However they can be assisted here if they have some idea for what purpose the user wishes to use the pictorial database. For example if the user wishes only to locate pictures which meet a certain criterion such as towns on a map which have specific spatial arrangements to each other then the designer can ignore the encoding of other objects attributes and concentrate on the encoding of the spatial position of the objects within the picture. This manual extraction of pictorial information for storage into a pictorial database management system is represented schematically in figure 6.11.

To manually encode information from a picture there must also be some understanding on how the human brain processes pictures. In principle there are an infinite number of interpretations of any picture, however many of these interpretations are wildly implausible (Goodman 1968). Kosslyn and Chabris (1990) have studied a subject's ability to categorise and name pictures and their contents and have concluded that pictures are named with remarkable consistency and that several factors influence the names assigned. These factors are:

6.3.5.1 Typicality

Input shapes are matched to similar shapes stored in visual memory (Jolicoeur, Gluck and Kosslyn 1984). If a previously seen object is very distinctive, it is likely to be stored as a separate representation in memory because it is distinctive, when first seen, it did not closely match any other stored representation and hence a new representation was created. In contrast, if an object is typical of a category, it is likely not to be stored as a separate example since its image already corresponds to at least one stored representation and hence a new representation need not be created when it is seen. Thus a typical shape will tend to match a stored visual representation that will later be categorised by its category name. Atypical shapes in contrast, may match a representation of an example better than the representation of the typical shape.Therefore more typical examples of a category are named as members of the category more often and more quickly than less typical members. For example a picture of a canary is named 'bird' more frequently and rapidly than is a picture of a swan.

FIGURE 6.11



MANUAL TECHNIQUES USED TO ENCODE OBJECT POSITIONS

6.3.5.2 Level of Hierarchy

Subjects tend to name an object at the 'basic level' of the object's taxonomic classification system, the basic level being the most inclusive level at which the members of a category have very similar shapes (Rosch, Mervis, Gray, Johnson and Boyes-Braem 1976). For example a Cox's apple would be named 'apple' (the basic level) rather than 'Cox's apple' (the subordinate level) or 'fruit' (the superordinate level). All Cox's apples have very similar shapes, but the shapes of apples in general overlap only slightly less than those of Cox's apples; in contrast, shapes of members of the category fruit (which include bananas, grapes, oranges etc) do not overlap very much. Thus the basic level is apple.

6.3.5.3 Interactions between typicality and level of hierarchy

The name people assign to a picture depends on both its typicality and its level of hierarchy. Holicoeur, Gluck and Kosslyn (1984) concluded that the more atypical an object is for its category, the more likely it is to be named at a level subordinate to the category description. In addition if an object is a typical example of its basic level category it is assigned a basic-level name more quickly than are atypical objects in the category. However if the object is an an atypical example, it is named more often not at the basic level, but at a level subordinate to it. Thus for example, a penguin is named 'penguin' more often than it is named 'bird' when subjects are given alternative labels in advance and asked to name the object shown.

6.3.5.4 Conclusions about naming objects

Kosslyn and Chabris (1990) concluded that for object's within a picture to be easily named then they should be drawn with intact contours, have as many distinctive parts visible as possible, have parts in their typical locations, and be presented upright. In addition if the picture is used to label a general category, and hence the name of the category should be evoked by it, then the pictured object should have a shape similar to other members of its basic-level category; in contrast if an exemplar name is required, the object should have a distinctive shape. The shape should be familiar, and the name of the object should occur often in the language. Within our HyperCard database system discussed in chapter 7, we have attempted to take account of these psychological conclusions about naming of objects where possible.

6.3.6 The advantages of querying pictorial databases visually rather than textually

These advantages include:

1) With regular users of a system, an iconic interface can offer improved performance in undertaking interactive tasks (Gittins 1986).

2) Match between representation and picture.
Visual representations of pictorial information more closely match the users memory of a picture than a textual description. This implies that the user selecting an appropriate visual representation is more likely to find the pictorial information that they require provided that there is a reasonable match between the representation and the actual picture

3) Difficulty in the textual classification of visual data.

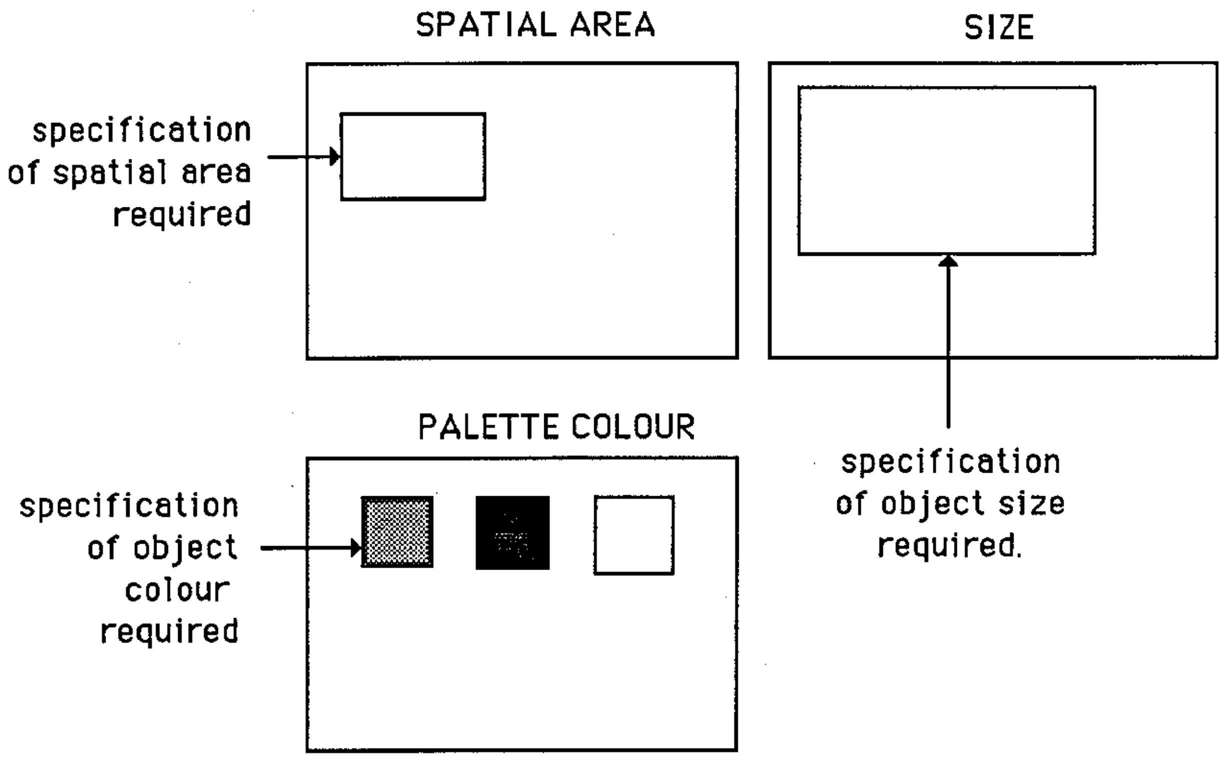
In order to compare user queries with pictorial information stored in a pictorial database then unless digital analysis techniques can be applied textual descriptions are required for the objects and their attributes within the pictures present. However users attempting to locate pictorial information based directly upon these textual descriptions may encounter problems since the concepts underlying the conversion of pictorial information into textual data may be unfamiliar to the user. For example only when users understands how size has been textually specified can they textually indicate the nature of their object size query. However when visually rather than textually choosing such information the conversion from visual query to textual data can be left for the database management system to determine rather than the user.

4) Problem of textual specification of data by the user.

Many visual queries are difficult to formulate textually. One problem here is that the object's visual appearance may only be vaguely remembered. For example a pictorial database query in which the user specifies that the picture required contains an object approximately in the centre of the screen, is of medium size, and is of a chosen colour, is a query difficult to specify textually because of the vagueness inherent in such a query. However in allowing a range to be visually specified by the user, (by allowing the user to draw a box representing the approximate size of the object of interest for the size attribute, and by selection from an on-screen palette in the case of colour) may help the user to formulate the possible object attribute ranges in their visual query. The database management system can then convert these visual query ranges into a textual or numerical form for comparison with the information stored within the knowledge base. Such a visual query is shown in figure 6.12.

Of course the wider the visual specification then the more probable it is that a greater number of matching pictures from the pictorial database will be displayed. However by comparing the objects shown on the screen with the textual representations described within the knowledge base users can be increasingly more specific in the nature of their visual query, finally locating the picture that they require.

FIGURE 6.12



VISUAL SPECIFICATION OF A QUERY RANGE

6.3.7 Problems with Visual Techniques of Picture Selection

One of the problems with visual selection of pictorial representations is that there are situations in which users know nothing about the appearance of the objects that they require but instead knows the name. For example the users may be only interested in particular kinds of object, say different kinds of cars, where although the name of the type of car is known, what this type of car looks like is not known to the users. For example some cars are difficult to distinguish except by name, (such as the physical differences between say, a Cortina Mark 1 and a Cortina mark 2). In this case although these names are known, physically being able to select the appropriate iconic representation of the car required may be difficult. Therefore if pictorial database users know only the names of the required objects and are not aware of their visual differences, then they will have difficulty in selecting appropriate icons from an iconic menu.

Another important problem with visual selection of object representations is to what degree can a generalised icon be said to match to an object present within a picture. For example if an user of a database management system were to select the icon of a cat from an iconic menu, should different pictures of the same cat be displayed to the user or only different cats? What cats constitute matching the users query? Should cats on the walls of Egyptian Pyramids be considered as matching the user's query? Should partial views of cats within pictures also be considered as matching such a query? Again the choice as to what objects in pictures match an icon is usually initially determined by a database designer when developing the database management system rather than the user when working with it. The problem for the database designer here is what objects can be considered to belong to a chosen class and therefore as matching to a chosen generalised icon? The success of the users in obtaining the information they require from a pictorial database will depend in part of the closeness of the user's perception of the object to the available icons within a menu, and in part on the closeness of the iconic representations encoded by the designer to the actual objects present within the pictorial database.

6.3.8 Conclusions about Selection of Object Representations

Representations of an object can be either in the form of generalised or specific icons. Generalised icons represent a class of object whereas specific icons are specific to an object within a pictorial database. These representations of objects can be matched to either pictures or topological representations of pictures. There are however problems with the descriptors used to match the representations of objects to the objects themselves. These include how the descriptors are textually described and which visual attributes should be chosen.

6.4 THE SPATIAL ARRANGEMENTS OF OBJECTS

6.4.1 When is spatial analysis likely to be useful?

It is suggested that the spatial analysis of pictorial information might be useful under the following conditions:-

Firstly there should be a 'reasonable' number of objects contained within a picture enabling spatial attributes of objects to be analysed. Pictures which have only a small number of objects have consequently less spatial information upon which spatial search techniques can be applied. However what constitutes a 'reasonable number of objects' depends to a large extent on the nature of the database. The greater the number of similar objects present the greater is the importance given to the spatial arrangements among these objects. For example for a database containing pictures of a large number of stars then the determination of spatial arrangements are vital in being able to specify collections of stars since other attributes (such as a star brightness or colour) may be the same for many stars. If a picture were to contain only a few stars then other attributes such as colour or brightness may be sufficient to distinguish them.

Secondly when considering which application area to apply spatial analysis to, then consideration of the spatial distance between objects should be taken into account. Pictorial databases which contain pictures of similar objects bunched closely together are more difficult to remember spatially because the bunching rather than the spatial arrangements between objects tends to be remembered (the Gestalt effect).

Thirdly the objects should be relatively simple. Complex objects which have many different features and attributes are considerably harder to analyse spatially. These objects often require that other visual attributes such as colour, size, shape and orientation be taken into consideration in locating a picture. The descriptions of such attributes thereby places more emphasis on the appearance of the objects and their visual attributes rather than their spatial locations or relative spatial positions with respect to other objects. Therefore it is suggested that a purely spatial technique of identifying pictures is of greatest importance in pictorial databases where the objects within the pictures are relatively simple.

Fourthly spatial searches should be applied to principally two-dimensional data within a pictorial database. Three-dimensional information represented two dimensionally in a picture is harder to analyse spatially because of the added complexity that object information may be hidden by other objects. Under such conditions users make assumptions about the nature of hidden objects which may not be correct.

To conclude, in pictorial databases in which there is a greater variety of different objects, it is suggested that the spatial information associated with an object is but one of the attributes upon which visual searches can be based. Under these circumstances other attributes such as name, size, shape and orientation become increasingly important in the identification of pictorial information and it is suggested that these attributes can also be used to find specific pictures. A pictorial database management system that visually depicts both spatial information and other visual attributes is demonstrated both in the next chapter and appendix 4 where a system has been constructed which allows the user to locate many of the paintings of Vincent Van Gogh by depicting object attribute values as visual queries.

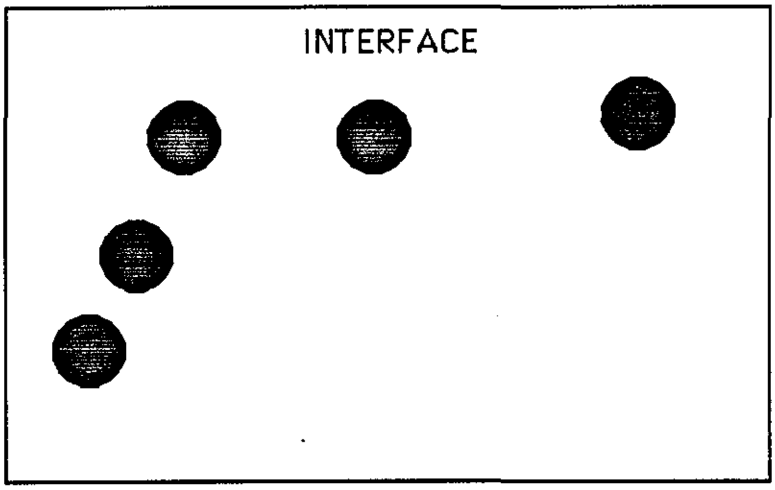
6.4.2 Specification of a Visual Query

Specification of a visual query consists of choosing appropriate objects representations and depicting visual attributes such as the spatial information associated with the objects within a picture. It has been suggested that this visual depiction both provides a more detailed query of the pictorial database and a closer match to the subject's visual representation of the picture required. One of the major advantages of such a system is that it takes into account the nature of the way that people remember spatial relationships between objects since it has been shown that people subconsciously remember the spatial arrangement of objects arranged within an information space (Deregowski 1974).

Where pictures within a pictorial database contain an equal number of identical objects differing only in the way that these objects are spatially arranged, then it is suggested that a visual query can be based upon the object's spatial locations and/or spatial arrangements with other objects. For example the picture below contains five identical objects arranged within a rectangular boundary. In order to select this picture from a pictorial database containing pictures of identical objects with different spatial arrangements, then one possibility is to allow the user to depict the locations of the objects by placing representations of them on the user interface with a pointing device such as a mouse. Having selected the representations and placed them on the user-interface in what the user feels are the correct spatial locations to locate a matching picture, then the closest matching picture to the user's depicted query can be determined by the database management system and displayed to the user. In the case of figure 6.13 the database management system would display image number one from figure 6.14, since the locations of the objects within the interface shown most closely visually correspond to that picture from the six shown.

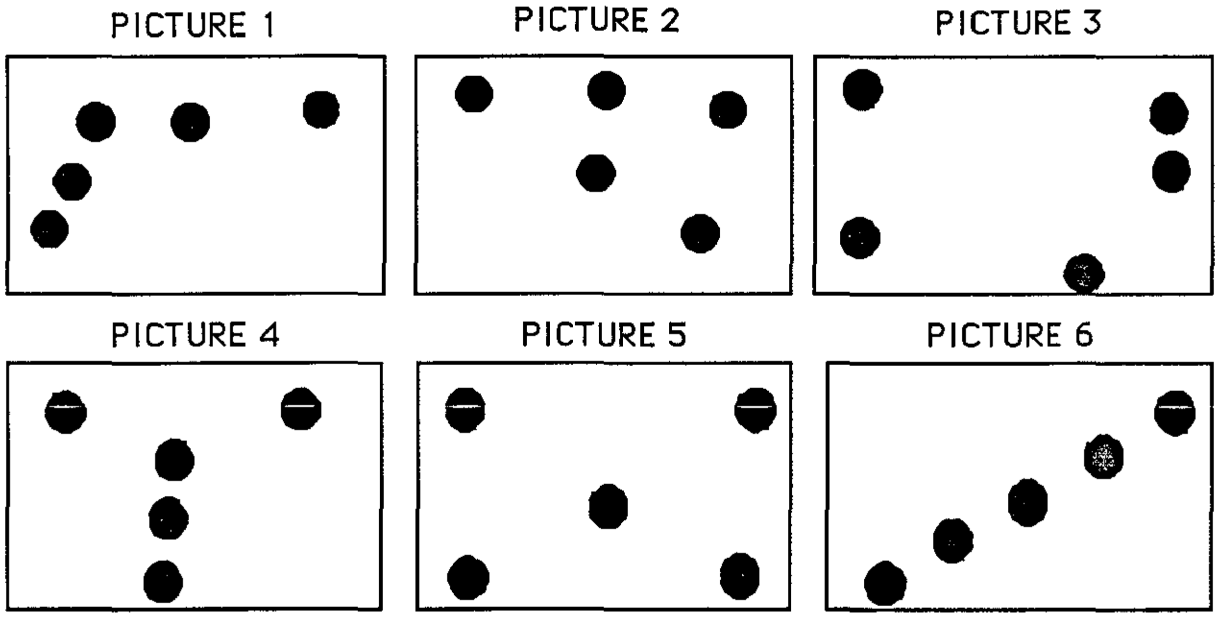
This depicted query allows subjects specifying the query to use their memory of the locations of the objects to locate the picture from a pictorial database. One of the strengths of selecting pictures by visually depicting objects on an interface is that this depictive technique can be used in the identification of pictures which have memorable spatial arrangements.

FIGURE 6.13



IDENTICAL OBJECTS SPATIALLY POSITIONED ON A INTERFACE

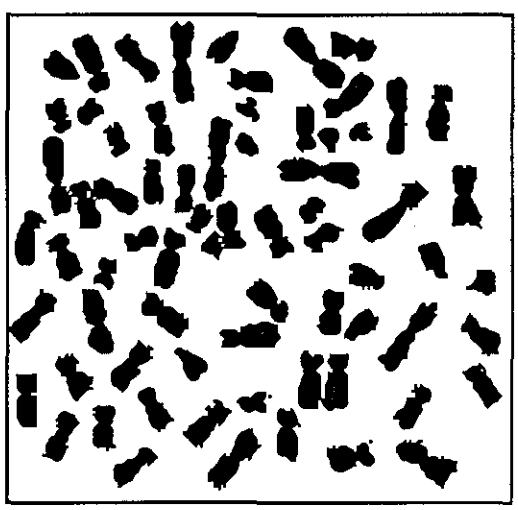
FIGURE 6.14



DATABASE OF PICTURES CONTAINING EQUAL NUMBERS OF IDENTICAL OBJECTS

A practical application of this principle of identifying pictures based upon the spatial arrangements of objects within them is in the analysis of a pictorial database of biological chromosomes, which is illustrated below.

FIGURE 6.15



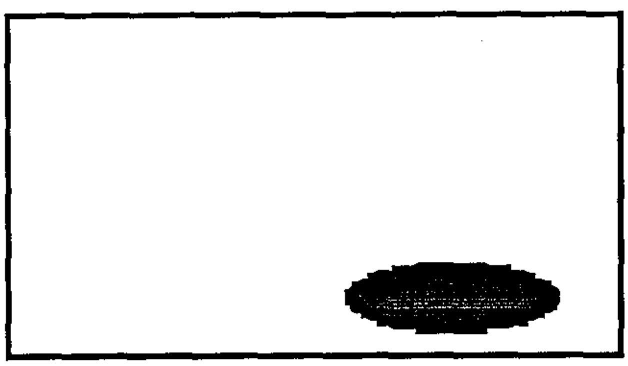
SPATIAL ARRANGEMENTS OF CHROMOSOMES WITHIN A PICTURE

Here the basic objects (chromosomes) are all visually similar, therefore different pictures of such chromosomes might be distinguished by the way the chromosomes are spatially arranged within a picture. In many cases the exact nature of an object (or objects) present within a picture may not be important, however the spatial locations or relationships between objects may be of considerable importance both in the identification of pictures and as information in its own right.

For example in the case above, particular spatial arrangements of chromosomes are important in medical diagnosis (Mckusick and Clairborne 1973). Therefore a query might be constructed by placing representations of chromosomes on an interactive user-interface and manipulating these representations into the desired spatial configurations after which the database management system matches the positions of these chromosome representations to pictures of the actual chromosomes stored within a database.

Similarly the same depiction technique can be applied when the objects in a picture are 'real' objects rather than simple shapes. Therefore the following elliptical shape on the user-interface:

FIGURE 6.16

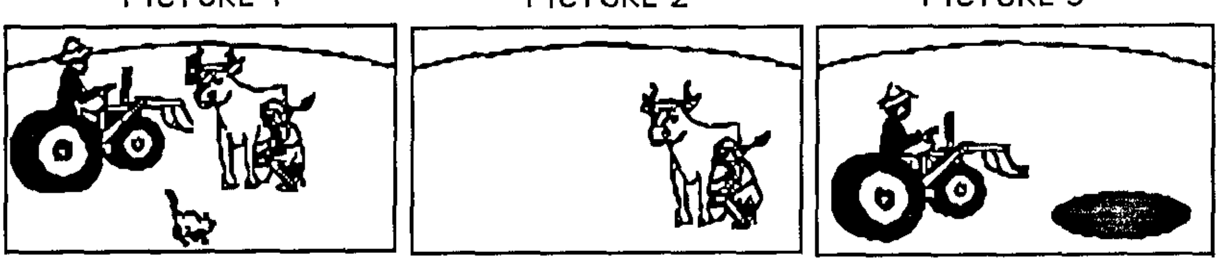


ELLIPSE DEPICTED ON USER INTERFACE

could be used to select PICTURE 3 from the choice below.

FIGURE 6.17

PICTURE 1 PICTURE 2 PICTURE 3

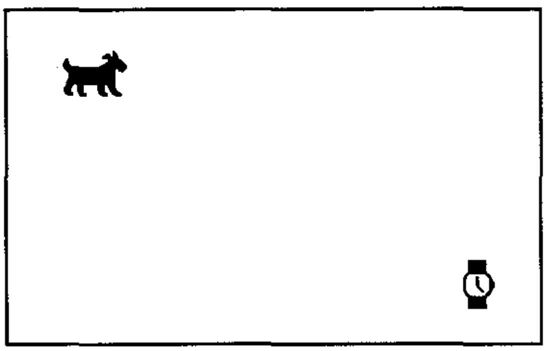


SIMPLE PICTORIAL DATABASE OF THREE PICTURES

Here again the visual query is based on the ellipse's object representation and spatial location within the picture rather than the physical descriptions or attributes of the other objects present. It is therefore suggested that one method by which 'real-world' pictures can be identified from a pictorial database is by the visual querying technique of choosing an appropriate icon from an iconic menu and then placing it on the user-interface in a position which matches to its position within a picture. Pictures with matching specified visual queries are then displayed to the user.

Thus for example in figure 6.18 the user depicts a visual query in which a dog is located in the top left hand corner of a picture and a watch is located in the bottom right hand corner of the picture required.

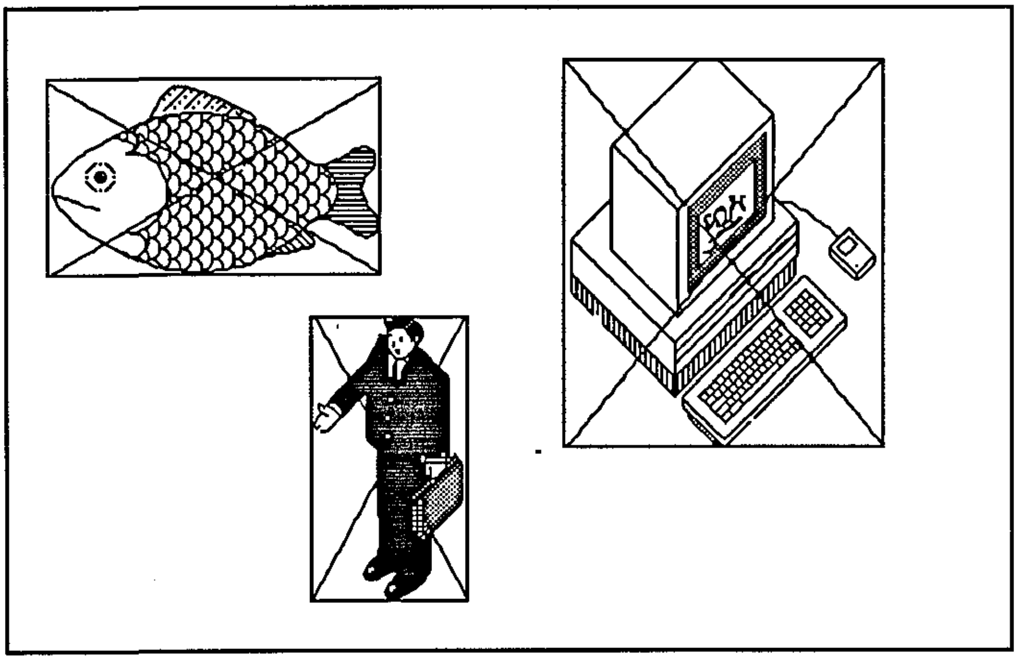
FIGURE 6.18



DEPICTED VISUAL QUERY

In order to specify the positional coordinates of both the iconic representations and the objects within the corresponding pictures, one technique that can be applied is the bounding box method in which an object (what ever its size, shape or appearance) is enclosed within a bounding box (Chang, Yan, Dimitroff and Arndt 1988). It is suggested that the centre of this bounding box can thereafter be used as a point representing the object's position. Object size can also be specified by the size of the bounding box. This is demonstrated in figure 6.19. In this figure the objects are contained within a picture. To select this picture matching icons can be selected from an iconic menu and placed in the correct positions on the user-interface.

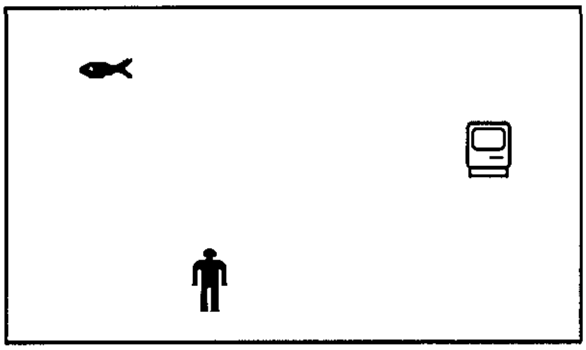
FIGURE 6.19



BOUNDING BOX TECHNIQUE APPLIED TO THREE OBJECTS

This placement of possible corresponding icons on an interface is shown in figure 6.20.

FIGURE 6.20

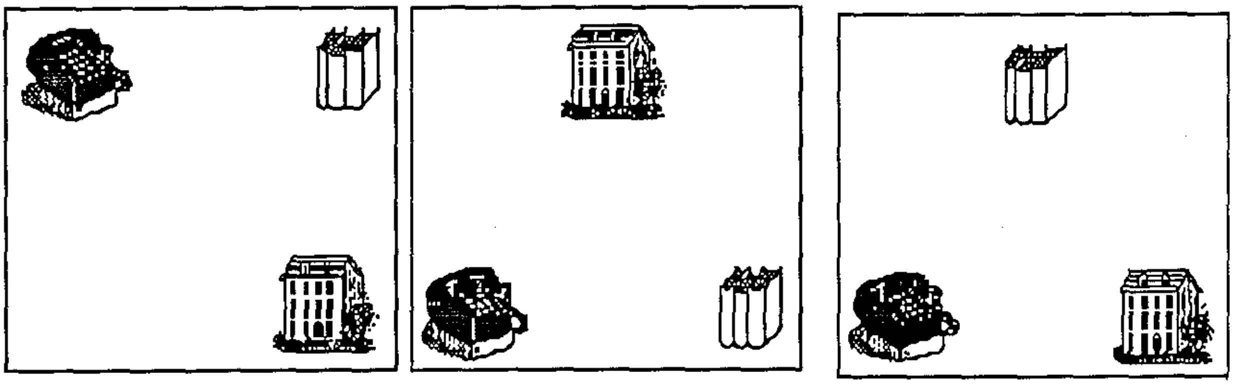


ICONIC INTERFACE CONTAINING ICONS

After placement of the icons, a database management system then attempts to match the nature of the icons (what they represent) and the coordinates of the centres of the icons bounding boxes (where they are), to the objects within the database. Pictures which match this visual query are then displayed.

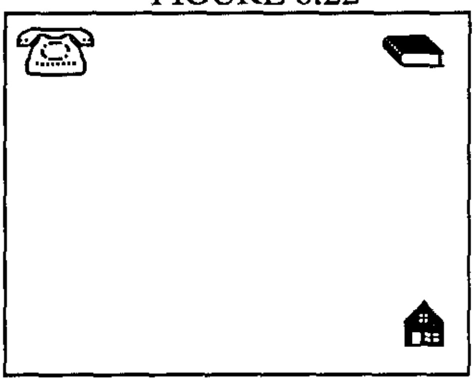
This technique of placing iconic representations in spatially significant positions is particularly applicable in locating a picture from pictures containing the same objects but in different spatial positions. For example in the case below, three pictures contain a telephone, a book and a house. One of these three pictures can be selected from the selection by depicting iconic representations in appropriate matching positions to the positions of the objects in the picture required.

FIGURE 6.21



THREE PICTURES ALL CONTAINING A TELEPHONE, A BOOK AND A HOUSE

FIGURE 6.22

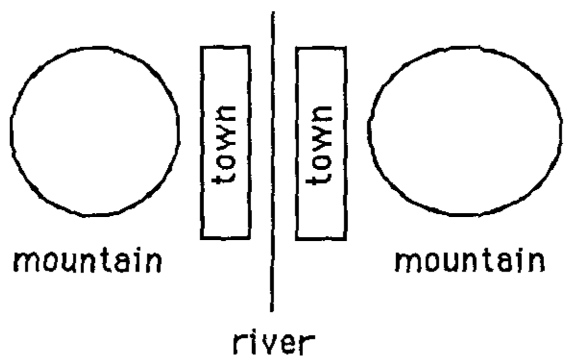


THE ICONIC QUERY (MATCHING TO THE FIRST PICTURE)

Since the iconic query above most closely matches the first picture, that is the one that is displayed to the user as matching the depicted visual query. Using this technique quite complex spatial relationships between objects can be searched for, providing that the objects in question have pictorial representations (icons) which the user can select. In general the ability to spatially position objects is developed independent of any experience of computers (Deregowski 1974). Children at a very early age learn to perceive and understand the nature of the world around them, and with this comes the ability to perceive and spatially manipulate objects in their surroundings. It is suggested that in many cases a system which used such a depiction technique to locate pictures would therefore be easy to use even for users who have little experience of computers since there is no textual programming language to learn, only the ability to visually select icons matching to objects within pictures and to specify the objects spatial locations by positioning the icons on an interface.

In the geographical domain such spatial analysis may have many applications. For example, such an analysis could be used in locating specific spatial arrangements of geographical features. In the figure below an iconic representation of two towns with mountains on either side and a river between them is shown. Spatially placing these representations on an interface could conceivably call from a geographical picture database all geographical areas within the database whose objects meet this spatial arrangement of icons.

FIGURE 6.23



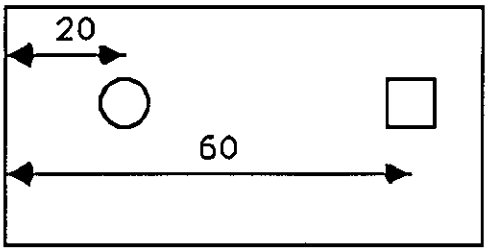
APPROPRIATE ICONS WITH THE REQUIRED SPATIAL CONFIGURATIONS FROM WHICH MATCHING INFORMATION CAN BE LOCATED FROM A PICTORIAL DATABASE

When spatially locating pictures based upon their object information, it is suggested that two basic forms of spatial analysis can be applied, which we have named absolute and relative spatial coordinates.

6.4.2.1 Absolute Spatial Coordinates

Absolute spatial coordinates specify the positional location of an object within a picture. This is shown in figure 6.24.

FIGURE 6.24



ABSOLUTE SPATIAL ANALYSIS

For example the user may remember two objects within a picture one of which was located in the top right hand corner of the picture while the other was located in the bottom left hand corner of the picture. Here then it is the absolute location of the objects within a picture that is important rather than the spatial relationships between them. These absolute spatial coordinates can in general be applied to two types of pictorial data.

Case 1) Pictures in which the objects differ visually and the absolute position of objects can be used for picture identification.

Case 2) Pictures in which objects are identical and the absolute position of objects can be used for picture identification.

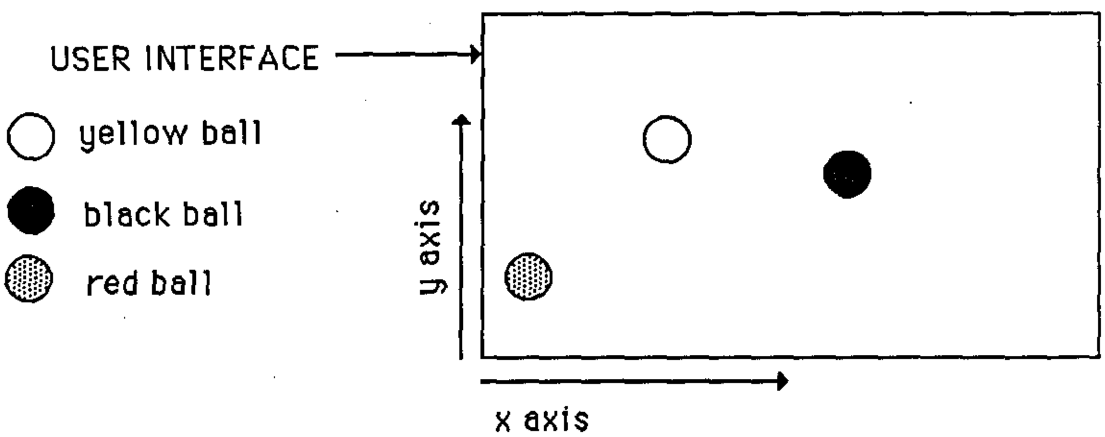
6.4.2.1.1 Case 1

Here the objects within a picture can be considered to be restricted within a boundary. This boundary constitutes the Cartesian axis from which the positions of the objects are specified. Objects within such pictures differ in one or more of their physical attributes such as colour, size or orientation. A combination of these physical attributes combined with an object's location within a picture can then be used to identify specific pictures. For example in appendix 3 a pictorial database has been constructed which consists of pictures of a billiard table with various positions of different coloured billiard balls on it.

Each picture within this database contains the billiard table and three balls having colours yellow, red and black. It is this attribute of ball colour that enables user specified queries of the ball's positions to be compared. Thus a query as to the position of the yellow ball within the user-interface is compared with the actual position of the yellow ball within pictures contained in the pictorial database. Similar comparisons can be carried out between the user-interface and pictures within the pictorial database for both the red and black balls. Since all the balls are restricted within the billiard table's cushions absolute positions of the balls within the query can then be directly compared with ball positions within pictures.

For example if the user directly manipulates icons of the three balls into the spatial positions shown on an user-interface in figure 6.25:

FIGURE 6.25



MANIPULATION OF DIFFERENT ICONS INTO APPROPRIATE SPATIAL POSITIONS

Then the coordinates of the balls on the interface are:-

red	x 10	y 20
yellow	x 50	y 80
black	x 100	y 70

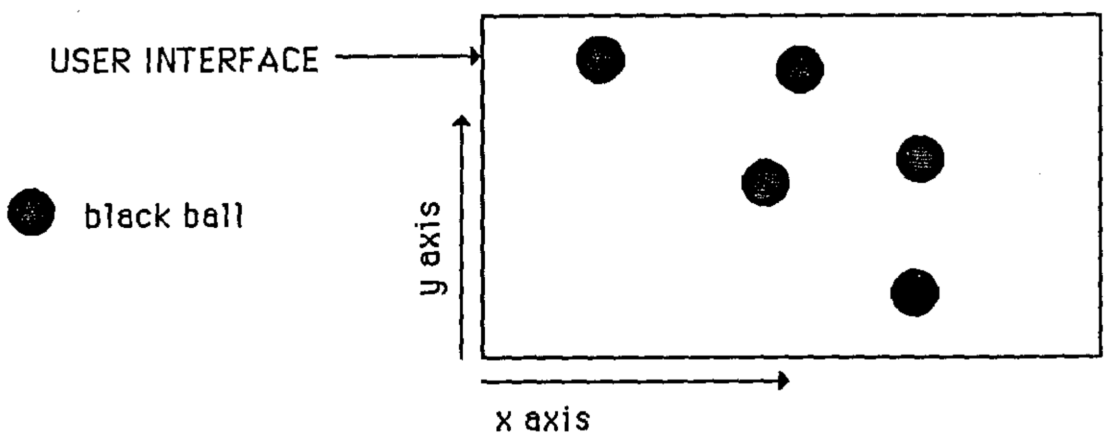
The ball positions within the user interface can now be compared with the actual positions of the corresponding balls within each of the pictures in the pictorial database in order to find the closest matching picture to the user's spatial query.

The closest match between the user's query and pictures within the pictorial database can be considered to be the one with the minimum mean absolute deviation between the balls on the interface and the corresponding colour balls in a picture (Naveh-Benjamin 1988). The picture from the pictorial database with the minimum mean absolute deviation between the user's query and pictures from the database can then be recalled and displayed to the users as the closest matching picture to their spatial query.

6.4.2.1.1 Case 2

This can be demonstrated again using the billiard table example discussed above, but which instead has on the table billiard balls of only one colour. Under these circumstances the user specified query to find a picture containing a specific spatial arrangement of billiard balls would entail that the position of each ball on the interface be compared with the positions of balls within a picture in order to find the closest matching configuration. This ball position query will need to be done for all pictures. Thus in the case below

FIGURE 6.26



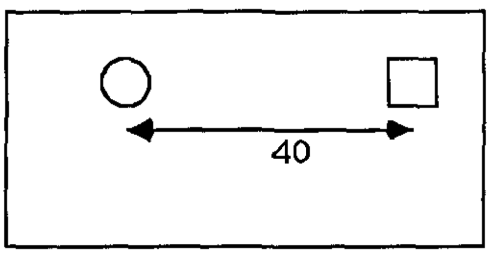
MANIPULATION OF IDENTICAL ICONS INTO APPROPRIATE SPATIAL POSITIONS

each ball position within the interface needs to be compared with all ball positions within a picture to find the closest matching ball. The distance apart between balls specified within the user-interface and the corresponding balls within a picture contained in the database can be calculated using pythageros theoem since the absolute locations of the balls within the interface and the balls within the database are known. When all the most closely matching balls to the user spatial query have been calculated within one picture, then a similar procedure is carried out for each picture. After this, the closest matching arrangements of balls within each picture can then be compared with each other by using the mean absolute deviation in order to find the best overall matching picture to the user's spatial query. This best overall matching picture is then displayed to the user as being the picture that most closely matches the ball configuration specified on the user-interface. In this way the user is able to locate a picture based upon the analogical depiction of identical objects on the user interface.

6.4.2.2 Relative Spatial Coordinates

Relative spatial coordinates use the distances between objects contained within a picture as a means of identifying them from a pictorial database. This is shown in the figure below.

FIGURE 6.27



RELATIVE SPATIAL COORDINATES

Using relative spatial coordinates, the objects to be identified can be grouped anywhere within a picture. For example the user may remember a picture in a database that contained three objects which appeared to form a triangle of a visually distinctive shape, although where this perceived triangular shape was within the picture could not be recalled. Here then it is the spatial arrangements between objects that is important rather than their absolute spatial location within a picture. Again as with absolute spatial coordinates, relative spatial coordinates can be subdivided into two forms.

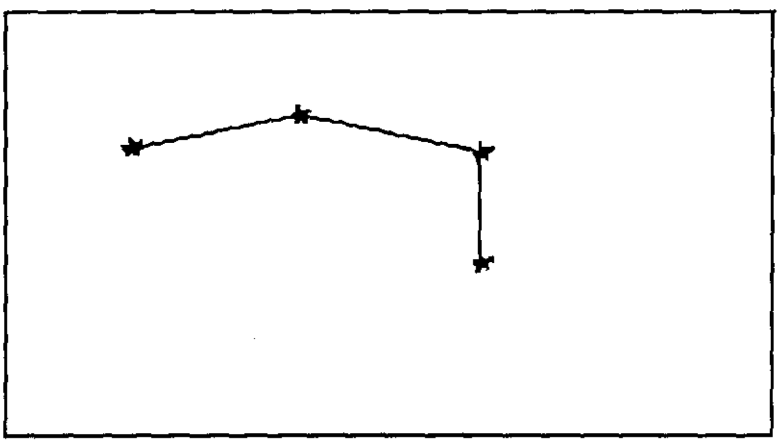
Case 1) Pictures in which all objects are identical and the relative spatial arrangements between objects can be used for picture identification.

Case 2) Pictures in which objects differ visually and the relative spatial arrangements between objects can be used for picture identification.

6.4.2.2.1 Case 1

An example of this case would be a pictorial database of stars in which although objects (in this case stars) within a picture can be considered to be identical, there is no boundary from which the absolute positions of the object representations within an interface can be directly compared with objects within pictures. Therefore within such a system the user requesting a particular spatial arrangement of objects (in this case stars), requires comparisons of representations of the objects with the actual objects in any spatial position of each picture contained in the database. For example if users specify the query in figure 6.28 as being a spatial arrangement of stars that they remember from a picture somewhere in a pictorial database.

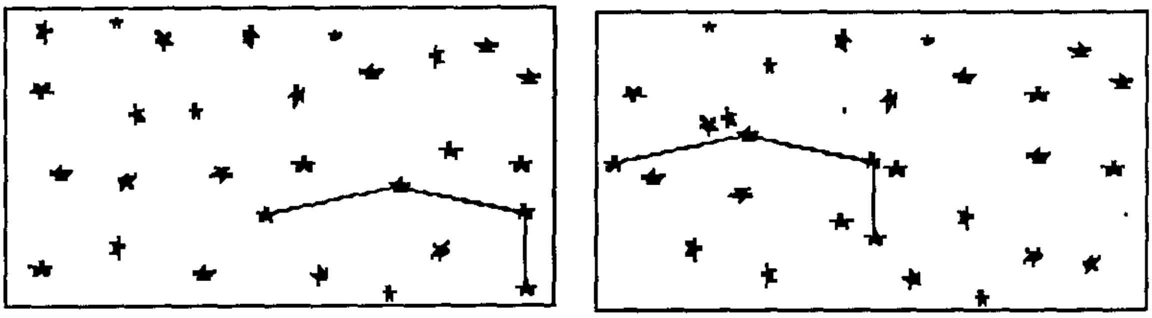
FIGURE 6.28



DEPICTING A STAR SPATIAL QUERY ON AN INTERACTIVE INTERFACE

Then matching spatial arrangements within two of the pictures within a pictorial database of stars might be:

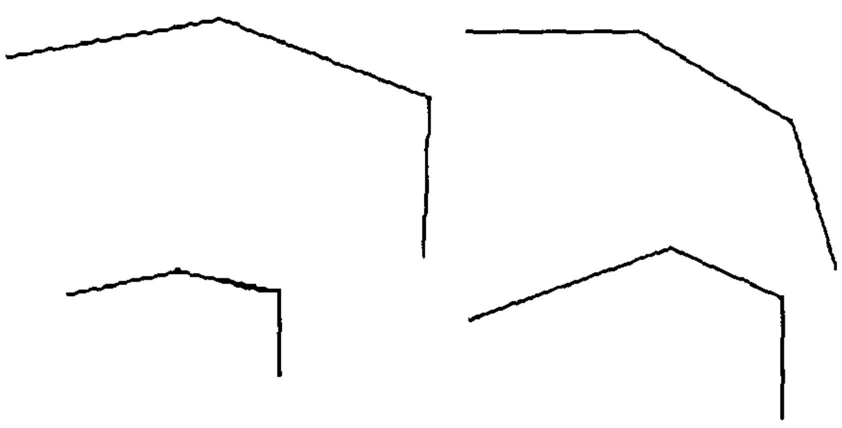
FIGURE 6.29



MATCHING SHAPES WITHIN TWO STAR MAP PICTURES

Here although the stars form a similar shape as that specified within the user interface, the absolute location of the shape within the pictures differs. Therefore a much more complex comparison between the user query and the pictorial database is required to identify the matching object's spatial arrangements within pictures. For example users specifying the spatial query in figure 6.28 may in fact not only require matching configurations from anywhere in a picture but also configurations which differ slightly in both angle, and distance from their query. Since users are unlikely to be able to specify exactly the relative spatial arrangement of objects that they require, tolerances must be allowed in both the object's distance apart and angle formed in order to locate matching pictures from a pictorial database. Therefore the user in specifying the spatial query above may also wish to find all the following spatial arrangements indicated below which differ slightly in both angle and size.

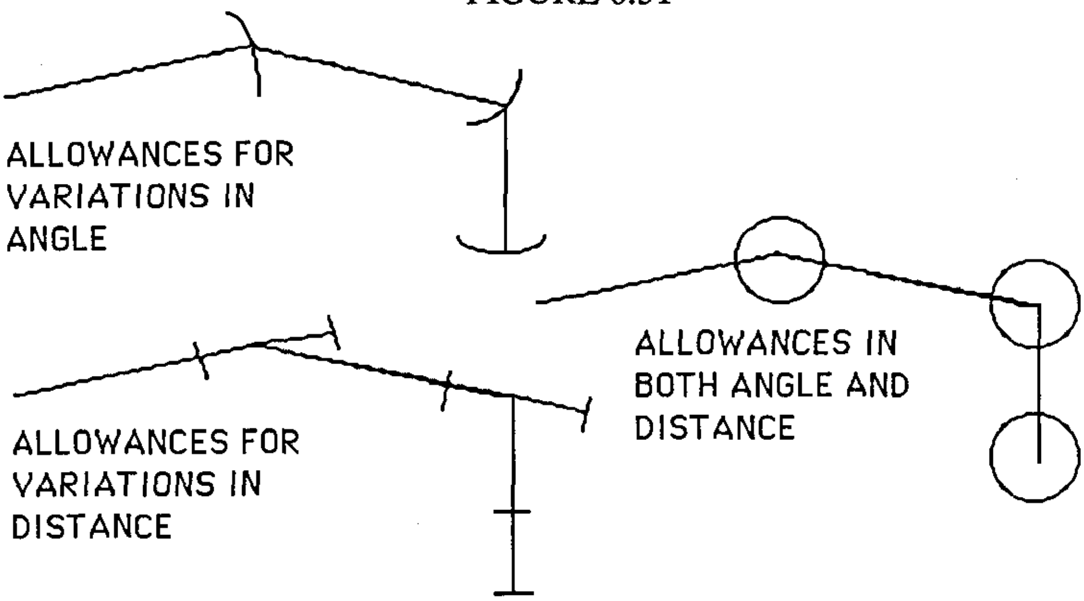
FIGURE 6.30



SPATIAL ARRANGEMENTS OF STARS THAT THE USER WISHES TO FIND USING THE ANALOGICAL QUERY ABOVE BUT WHICH DIFFER IN ANGLE OR SIZE FROM THEIR QUERY

These spatial arrangements of objects differ in both angle formed and distance apart, therefore in the initial query specification on the interface the degree of variability in both of these visual attributes needs to be allowed for. A possible solution to this problem is indicated in figure 6.31. Allowing for a 'circle of variation' around each object indicates to the database management system that objects that lie within this area can be considered as possible matches. However only where the complete set of objects lie within the spatial arrangement specified within the user-interface, can that set of objects be considered as matching to the user's query.

FIGURE 6.31



ALLOWANCES IN BOTH ANGLE AND DISTANCE APART OF OBJECTS WITHIN A USERS QUERY

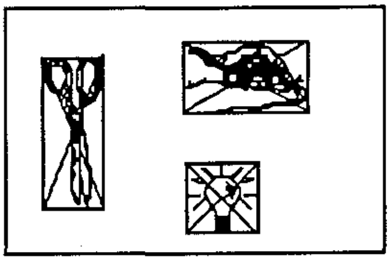
6.4.2.2.2 <u>Case 2</u>

Here the user relies upon the relative spatial arrangements of different shaped objects in order to identify a picture. For example the user may recall that three objects A,B,C within a picture had relative spatial arrangements of:- object A being above object B, object B being to the left of object C and object C being below object A. Here there is no means by which absolute positions of the iconic representations of objects within an interface can be directly compared with the absolute positions of the corresponding objects within a picture. Therefore a method is required that compares the relative positions of the user's spatial query with pictures from the database. A method suggested is Spearman's rank correlation coefficient. This is a formula used within statistics by which two sets of data can be directly compared where only the order of the items (not their absolute values) is known. The formula for its calculation is

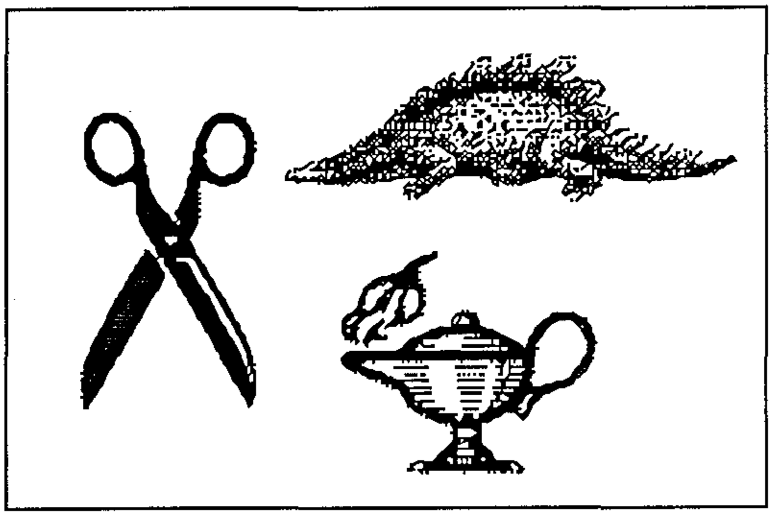
$$r' = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

where r' is the coefficient of rank correlations.
d is the difference between the rankings of the same item in each series and n is the number of ranks.
The use of this formula is demonstrated below. Here icons have been chosen which can be considered to match objects within pictures contained within a pictorial database. These icons can then be placed on the user-interface and matching objects and relative spatial configurations within pictures searched for within the database.

FIGURE 6.32



ICONIC REPRESENTATION ON USER-INTERFACE (POSITIONS OF BOUNDING BOXES SHOWN)



POSSIBLE MATCHING PICTURE FROM A PICTORIAL DATABASE

Thus in the case above the user is searching for pictures which contain a dinosaur, scissors and a source of light, with the relative spatial arrangement shown. Here the three icons within the user interface have their bounding box centre coordinates of (say)

scissors x 20 y 40
dinosaur x 50 y 60
light x 30 y 10

One possible corresponding picture which contains the same objects has bounding box centre coordinates of (say)

scissors x 30 y 42
dinosaur x 55 y 65
light x 45 y 5

Therefore to check whether the objects within a picture correspond to the same relative spatial arrangement of icons within the users query then Spearman's rank coefficient of correlation is applied in both the x and y axis. In the x axis ranking the icons produce

scissors 1
animal 3
light 2

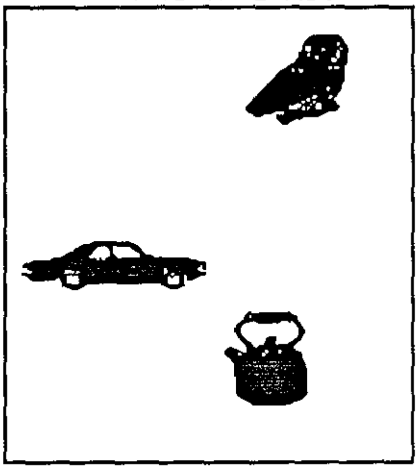
and in the case of the picture ranking the objects produces

scissors 1
animal 3
light 2

Entering these ranks into the Spearman's rank coefficient formula produces a correlation of 1. Similarly a rank coefficient of 1 is produced in the y direction. In cases where the ranks differ between icons on the user-interface and the corresponding objects in a picture, then the closeness of the rank correlation coefficient to 1 determines the degree to which the relative spatial arrangement of the iconic representations in the user-interface matches the relative spatial arrangement of the objects within the pictures.

Objects within pictures may then have similar relative spatial arrangements. For example in the iconic query shown in figure 6.33, icons have been selected representing classes of objects (KETTLE, BIRD and CAR) and the relative spatial arrangements required have been depicted.

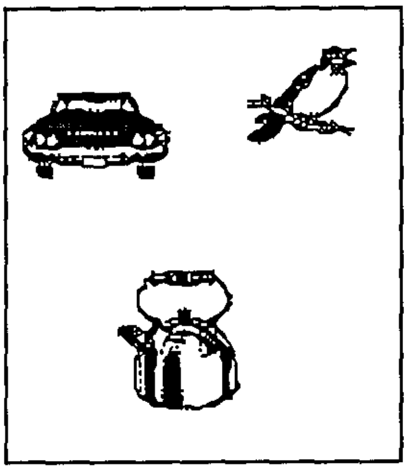
FIGURE 6.33



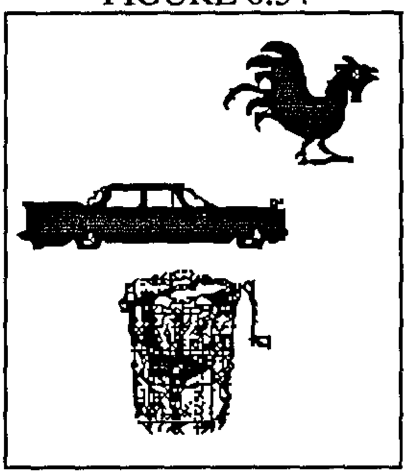
ICONIC QUERY DEPICTED ON USER-INTERFACE

This iconic query matches to the three pictures shown below containing non-identical objects within the same textual classes (car, bird and kettle). These triplets of objects also have the same relative spatial arrangements (car to the left of bird, kettle below car, etc).

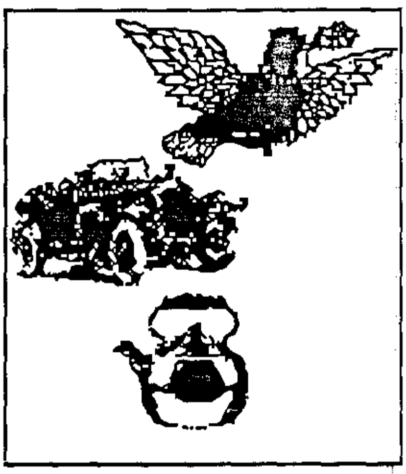
FIGURE 6.34



PICTURE 1



PICTURE 2



PICTURE 3

SIMPLE DATABASE OF PICTURES CONTAINING OBJECTS OF THE SAME CLASS

In order to select one of the pictures from the three shown above additional visual attributes are required which uniquely identify that picture. If one or more of the values of the object's visual attributes within a picture can be specified then this enables a more detailed search to be made from the three pictures shown. Such object attributes might include the colour, shape, size or orientation of objects within the picture. A detailed description of various possible visual attributes is given in appendix 2. In the above case PICTURE 1 has a car orientated differently from the car in PICTURE 3. Therefore if the value of the orientation attribute is depicted for the car, in combination with the depiction of the relative spatial arrangements then this should uniquely identify one of the three pictures from the choice shown. For the three objects present within the pictures in figure 6.34 the spatial relationships between them are:

Object CAR left-of Object BIRD

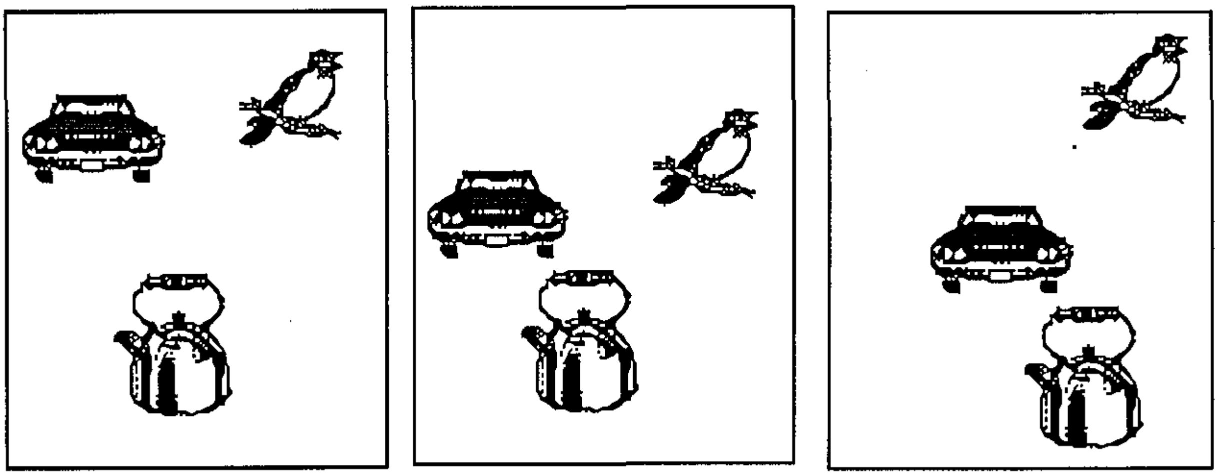
Object CAR above Object KETTLE

Object BIRD above Object CAR

Pictures within the pictorial database containing objects within the same class and having the same relative spatial arrangements may then be located and displayed based upon these simple relative spatial relationships queries. Additional visual attributes can be specified which distinguish between these pictures such as the object's colour or size.

However in order to differentiate between pictures containing identical objects arranged in identical relative spatial arrangements then additional spatial attributes such as the approximate absolute location of one or more of the objects within a picture are required. Under these circumstances the absolute position of an icon within the user-interface can be directly compared with the position of the corresponding object within the three pictures.

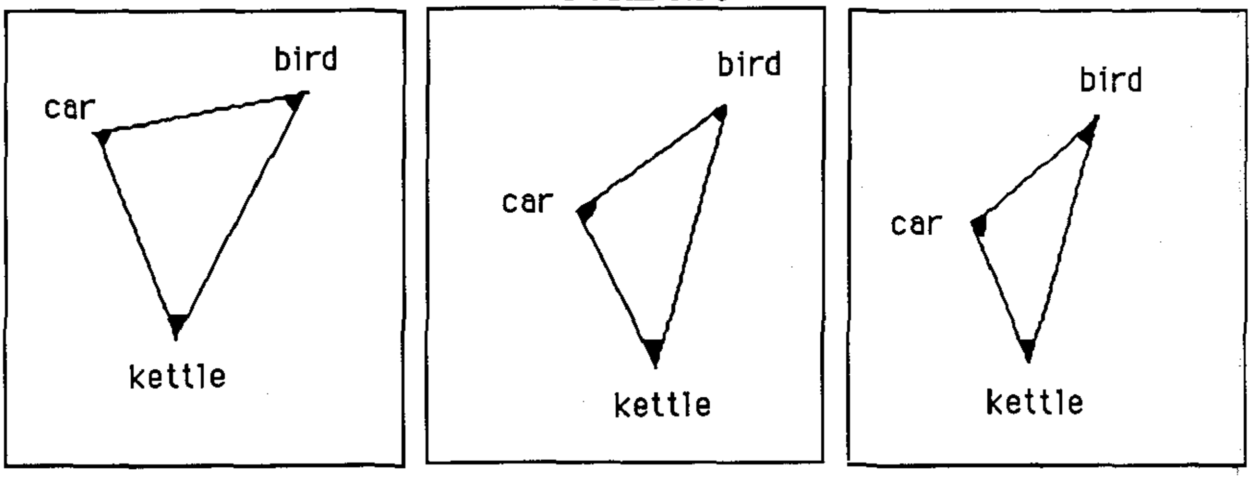
FIGURE 6.35



IDENTICAL OBJECTS IN SIMILAR RELATIVE SPATIAL ARRANGEMENTS

In the case of the three pictures above the corresponding angles, distances, sizes and shapes of the triangles formed are shown below.

FIGURE 6.36

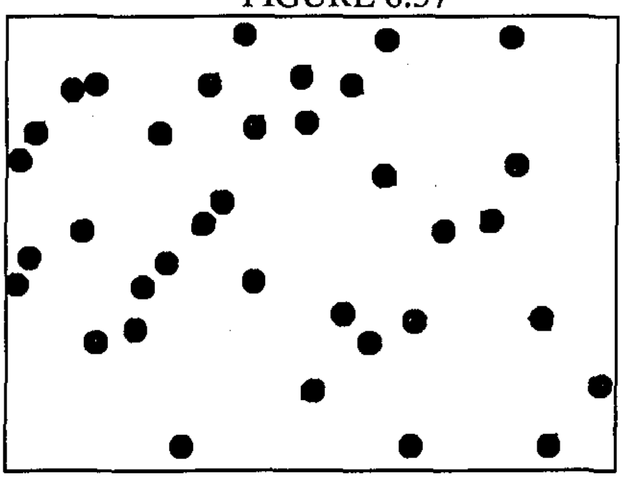


ANGLES AND DISTANCES BETWEEN OBJECTS WITHIN PICTURES

Specifying more detailed spatial attribute values such as the distance apart of the objects, their orientation to each other or the size of the triangle formed enables this information to be directly compared with corresponding spatial information extracted from pictures within the pictorial database, thereby locating the closest matching picture to the visual query. Thus for example the user knowing that the car is further away from the kettle than it is from the bird and specifying this as a visual query on the user-interface should enable the database management system to use this additional spatial information to locate a picture that the user requires. Examples of more detailed relative spatial queries are discussed below.

The picture shown in figure 6.37 contains many objects, each represented by the centre coordinates of their bounding boxes. If different attributes of the spatial arrangements of objects are depicted, then matching relative spatial arrangements of objects can be found from the picture.

FIGURE 6.37

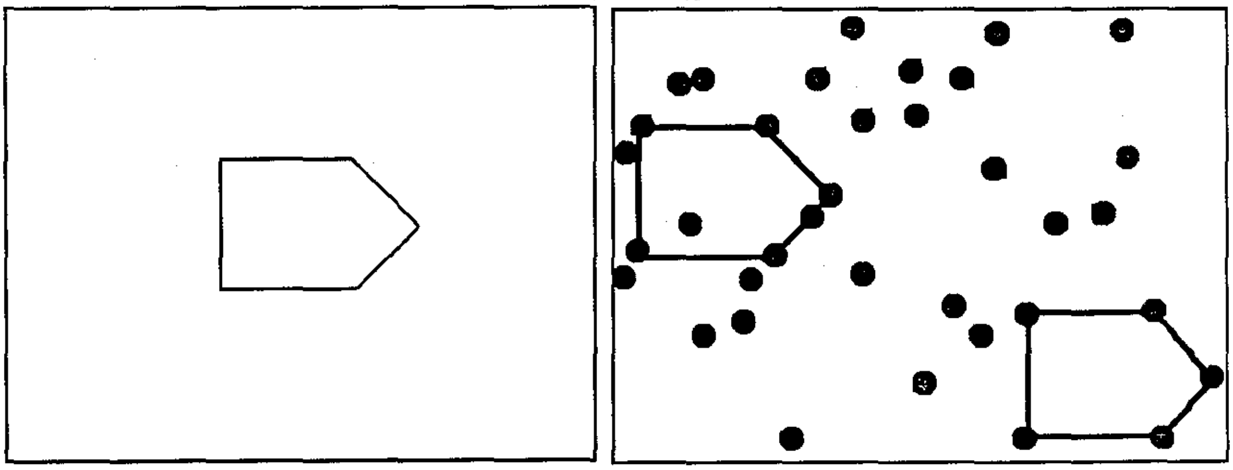


POINTS REPRESENTING BOUNDING BOXES CENTRES IN A PICTURE

For example possible spatial queries are shown in figure 6.38 in which spatial information has been depicted and matching information with variables in size, shape and orientation are shown.

Query One: relative position shape, size and orientation between objects fixed.

FIGURE 6.38

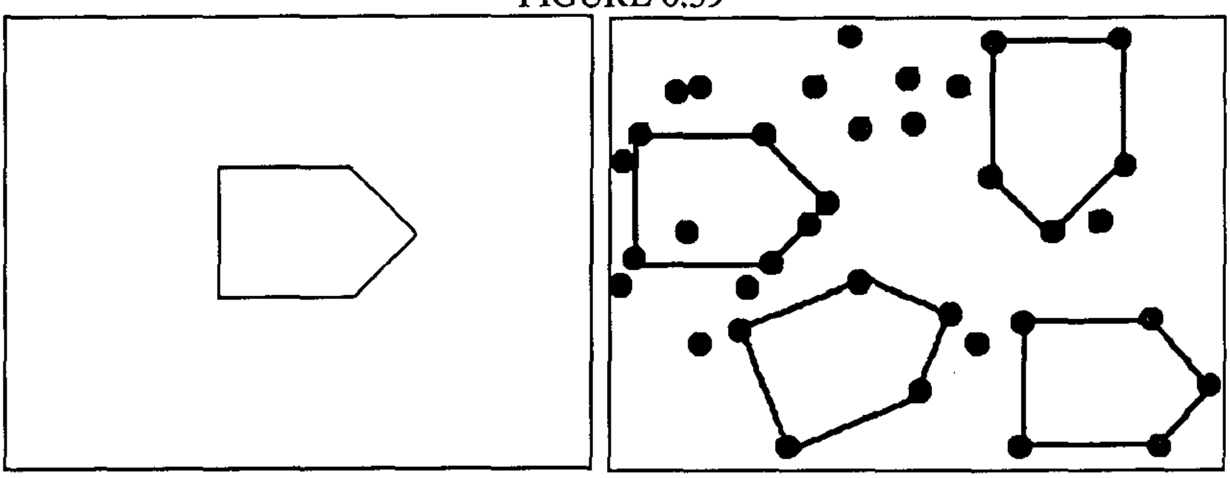


VISUAL QUERY SHAPE, SIZE, RELATIVE POSITION, ORIENTATION CONSTANT.

MATCHING INFORMATION IN PICTURE

Query Two: relative position, shape and size between objects fixed.

FIGURE 6.39

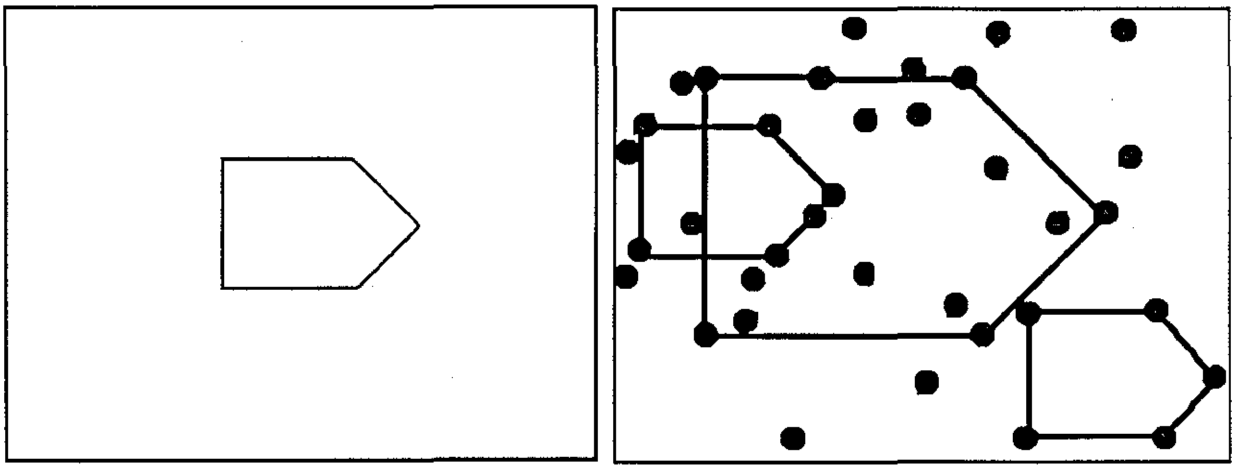


VISUAL QUERY SHAPE, SIZE AND RELATIVE POSITION CONSTANT

MATCHING INFORMATION IN PICTURE

Query Three: relative position, orientation and shape between objects fixed.

FIGURE 6.40

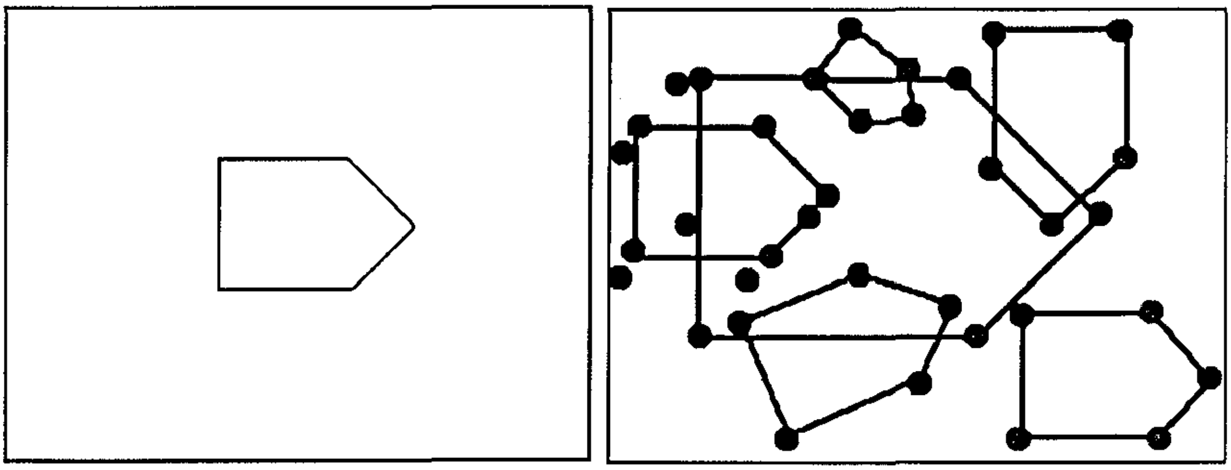


VISUAL QUERY SHAPE ORIENTATION AND RELATIVE POSITION CONSTANT

MATCHING INFORMATION

Query Four: relative position and shape only

FIGURE 6.41



VISUAL QUERY SHAPE AND RELATIVE POSITION CONSTANT

MATCHING INFORMATION

The combination of spatial attributes specified by the user therefore restricts the number of matching spatial patterns within the picture that are found in a visual query. Thus for example the user visually specifying values of the attributes of relative size, size, shape and orientation restrict the database management system to searching for a spatial arrangement of objects in a specific location, of specific size and of a specific orientation but which exist anywhere within a picture. Practically however combining a simple absolute or relative spatial query with other object attributes such as colour, shape and size should be sufficient to locate pictures without resorting to depicting detailed spatial arrangements between objects.

6.4.3 Conclusions about the Spatial Arrangements of Objects

Objects can be selected spatially either using an absolute or relative form of search. One possible technique of specifying such a visual query is by depicting the object's spatial information on the user-interface thereby allowing the database management system to locate matching pictures from a pictorial database. These spatial attributes can be combined with other object attributes such as size, colour and orientation for a more detailed pictorial query.

6.5 PRACTICAL REASONS FOR SELECTING PICTURES BASED UPON THEIR SPATIAL ARRANGEMENTS

These reasons include:

1) The user may require to see the exact position of the object specified within a picture in order for measurements to be taken from a printout. Thus by specifying the approximate location of an object on an interface, all pictorial database pictures which have identical objects in the same approximate spatial location can be recalled thereby enabling the user to identify a specific picture from which measurements can be directly made. Medical image analysis or map applications are possible examples.

2) The initial picture located may be the start of a motion sequence which the user wishes to see. For example on a videotape or videodisk the initial still frame may be followed by a sequence of motion information related to this initial

picture. Therefore in this case, locating the initial still frame by specifying the spatial location of objects within it enables users to find the picture and thereby run the motion sequence that they wish to see.

3) Pictures could be located within a pictorial database by indicating the spatial arrangements of objects present within them. Therefore this spatial attribute possibly in addition with other attributes of objects (such as colour, shape, size etc) could be used to locate a specific picture. All these practical reasons make use of a subject's ability to incidentally recall spatial information without conscious attention having being paid to it.

6.6 CONSTITUENTS OF SPATIAL INFORMATION SYSTEMS

As with all pictorial databases, databases containing pictures of objects and their corresponding spatial information require three basic components.

1) A database of pictures to be analysed, usually such a pictorial database is stored either digitally within a computer system or on a medium such as a videotape or videodisk.

2) A consultation facility which enables the user to query the pictorial database: in the case of pictorial databases which allow the specification of spatial queries, a study of spatial database systems (see chapter 4) suggests that these queries need to be interactive. This is because the underlying principle of a depicted query is that the user of the system should be able to manipulate visual representations of the object's thereby enabling users to relate their queries to their spatial memory of the objects locations and spatial arrangements to other objects. Therefore the ability to select, move and place icons within the user-interface is required. The database management system also needs to be able to detect the final positions of icons on the user-interface and compare the coordinates of these positions with positions of objects extracted from pictures stored in the pictorial database.

3) Data manipulation and storage which ensures that the query constructed by the user can be compared with objects contained in pictures (see chapter 3). In the

case of object's spatial positions then the coordinates of the bounding boxes of the object's iconic representations within the user's query can be directly compared with those of the positions of the objects within the pictures present in the pictorial database.In the next chapter a practical database management system is described incorporating these three components.

6.7 POSSIBLE APPLICATION AREAS FOR SPATIAL SEARCH TECHNIQUES

Possible application areas which might benefit by the application of spatial analysis techniques include such diverse examples as:-

Medical Image Analysis	Paintings	Archaeological Artifacts
Map Applications	Identikits	Military Applications
Cell Analysis		

or other information systems where the spatial organisation of objects is important. Applications such as identikits, archaeological artifacts and paintings contain objects whose relative distance apart are an important attribute. For example in the case of identikits the distance apart of two features (such as the nose and mouth) can be used to identify a particular face. Within other applications such as cell studies and military applications the determination of the spatial relationships between objects can indicate actions that should be taken, particular spatial arrangements of cells may indicate disease, whereas certain spatial arrangements of military units may for example indicate attack or defence formations within hostile armies. More detailed descriptions of some of the possible applications follows.

6.7.1 Military Applications

The spatial arrangement of information is important within the military map domain because particular arrangements of military vehicles and weaponry can be used as indicators of the intentions of the people deploying the equipment. Attack formations differ significantly from defence formations (Tranowski 1988). Additionally the spatial arrangements of selected military hardware might indicate a particular course of action is to be undertaken and therefore analysis of such data is important in the defence against such actions. The map source for such data usually comes from sources such as satellites.

6.7.2 Cells

Particular spatial arrangements between cells can in many cases indicate the presence or absence of disease. Therefore analysis of such information can help in the diagnosis of illness. If it was possible to digitally analyse the spatial arrangements of cells on a slide then this digitised information could be compared with the user's spatial cell query indicating a disease. All slides matching this user query could then be displayed for a more detailed analysis. In this way abnormalities in cell arrangements could be detected saving a great deal of time usually occupied by the manual analysis of slides.

6.7.3 Paintings

The problem of how to express space in pictures has always been an important one in art (Belhage 1989). Shapes and patterns within paintings are often characteristic of a particular artist. This may sometimes take the form of constant distances and angles between objects within paintings. Therefore these features can be used as characteristics of that particular artist and can be encoded and stored as such. The spatial arrangements of such objects can be recalled, analysed and compared with other artist's characteristic spatial arrangements of similar objects within paintings. Vaughan (1988) has used such a system in his Morelli system in which he attempts to establish an objective system for the recognition of the styles of different artists. Here he compares features common to all the objects he was classifying. Since he was working with artists who habitually depicted the human form, he chose to compare the ways in which they treated different parts of the body. Such a system becomes more effective the more pictures are entered into the pictorial database since the likelihood of detecting objects which closely match the user's specified spatial query between objects will thereby increase.

One particular feature that Vaughan paid attention to was the distinct ways in which different artists depicted ears. Different artists depicted ears differently not only in their size but also in the spatial arrangements between different parts of the ear. Morelli's system provided a universal method for identifying, classifying and comparing pictures. It attempted to use the analysis that digitised reproduction automatically performs on a picture to derive specific types of measurement. The result of this process was a set of numerical variables that can be compared. In effect it generates a 'unique identifier' for each picture by automated means. The system developed also recognised that different reproductions of the same picture refer to the same image and consequently that manual copies differed from originals. This is because manual copies however carefully made show variations of form that do not occur in photographic copies, however clumsily the latter are made. Another advantage of this 'Morelli technique' is that it provides a means of associating pictures that are formally similar to each other without actually confusing them with each other. This effect is such that a school of painters can even be detected by their depiction of similar physical objects (in the case above the physical object being ears). Therefore analysis of spatial distances within pictures allows closest matching objects to the user's spatial query to be detected.

6.8 CONCLUSIONS

Information contained within pictures can be located based upon spatial information associated with objects within them. This information can be found using an absolute or relative reference of search or a combination of both. Such a spatial query can be displayed analogically. An absolute reference of search relies upon being able to locate a picture based upon remembered object positions within the picture. A relative form of reference relies upon the objects remembered spatial configuration to surrounding objects such as object A is left of object B etc. This relative form of reference can be analogically specified either by placement of icons or depiction of the pattern formed by the objects. These spatial techniques allow users to identify pictures based upon their knowledge of how the objects are spatially arranged within a picture.

It is suggested that a generalised database management system which is able to encompass all manner of pictorial objects rather than those contained within a restricted application area, enables a wider variety of pictorial information to be analysed. Previous methods of locating pictures have relied upon textual methods such as keywords to find the pictures required from a pictorial database. However since keywords are fundamentally designed to access databases containing textual information, their usefulness in the case of pictorial databases is severely limited (see chapter 3).

In the analysis of pictorial data Abel (1988) has suggested that a recognition of the value of integrating image, coordinate and textual data is an important impetus to consider in the analysis of such information. It is also becoming increasingly recognised that effective utilisation of the visual system's spatial analysing capabilities can depend on exploiting its natural scene processing mechanisms. This suggests that object attribute values can be used for representing data variables within such scenes (Robertson 1988). A practical database management system has been built which attempts to take account of both Abel's suggestions and the conclusions of this chapter and this system is described in the next chapter.

CHAPTER 7

A PRACTICAL PICTORIAL DATABASE CLASSIFIED BY DIFFERENT VISUAL ATTRIBUTES

7.1 INTRODUCTION

In the previous chapter it has been suggested that both selecting visual representations and depicting the values of the physical attributes of objects (such as their spatial locations) allows the user to locate pictures from a pictorial database.

In what follows a practical database management system is described which allows the user to select a picture based in part upon visual attributes of the objects within it, one of these attributes being the spatial locations of the objects within the picture (Charles and Scrivener 1990 (see appendix 5)). Within this database management system, values of the spatial attribute can be selected either from a textual menu or by visually depicting its possible positions on a user-interface. Apple HyperCard software has been used to construct the database management system, combined with both a laserdisk as the picture source and an Oracle relational database for the storage of visual object attributes abstracted from the pictures.

7.2 VISUAL OBJECT ATTRIBUTES

Pictorial databases can be classified by many types of object attributes depending in part on the nature of the objects present. As detailed in appendix 2 there are several important attributes which can be used in the analysis of pictorial information. Psychological evidence such as Gestalt analysis (see chapter 4) suggests that the list below gives the most commonly used attributes in the description of an object. These attributes are:-

name
shape
size
position
colour
orientation
width
length

Thus a database management system which enables the user to locate pictures based upon values of combinations of these attributes is we feel a suitable method by which a picture can be located from within a pictorial database.

7.3 CHARACTERISTICS REQUIRED OF THE DATABASE MANAGEMENT SYSTEM

Canavesio and Marion (1989) have proposed that considering the human factors requirements of the users during the interface design improves its usability. In accordance with these human factor requirements and the conclusions drawn from the previous chapters, a database management system is proposed which has the following characteristics.

1) The ability to identify pictures without a detailed knowledge of the database structure (a requirement identified in chapter 4).

2) The ability to browse through the pictorial database, the browsing controls being easily reversible (a requirement identified in chapter 4).

3) The ability to pan and zoom information thereby allowing the user to move around the system both quickly and efficiently (a requirement identified in chapter 4).

4) The ability to easily update and add extra information to the system (a requirement identified in chapter 4).

5) The ability to become increasingly more specific in a visual or textual query in keeping with how subjects mentally describe objects (a requirement identified in chapter 5). Such a hierarchical selection system should also allow the user to browse the database by enabling each of the actions in the progressively more detailed selection to be easily reversible.

6) The ability to select iconic representations of objects contained within the pictorial database (a requirement identified in chapter 6).

7) The ability to select pictures based upon the spatial positions of chosen objects within the picture. Object representations (in the form of icons) can be chosen and a range of spatial positions for these objects within a picture can then be specified (see chapter 6). It is suggested that high resolution graphics should be used for a more 'life-like' representation.

8) The ability to visually depict object attribute queries on a graphical user-interface using a mouse (see appendix 1). It has been hypothesised (Guastello, Traut and Korienek 1989) that pictorial displays for an object-oriented vocabulary (in which objects can be selected) are easy to recognise and manipulate.

9) The ability to select pictures based upon incomplete information (Batley 1988). This allows the user to recall only a few values of the attributes of an object within a picture and still be able to find the required picture.

10) A carefully designed user-interface - the emphasis should be on ease of use so that the system is accessible to both naive and expert users (Batley 1988).

11) The ability to select information from a pictorial database using a variety of techniques (Abel 1988). Here it is suggested that some attempt should be made to individualise searching, for example by allowing users to choose between visual depiction and typing in commands.

In order to practically implement these proposals the database management system we have developed allows the coexistence of both iconic and lexical table representations. These representations allow the user to communicate with the database management system by entering queries in several different ways.

7.4 TYPES OF QUERY

The database management system developed within Apple's HyperCard was designed to handle four basic forms of query:-

1) Selection by icon from an iconic menu.

Here the iconic representation of an object can be selected from within an iconic menu. Having selected the appropriate icon further details about the object can then be obtained by combining this object representation with other search techniques. Generalised icons were used as object representations because of the inability to 'import' appropriate graphics representing specific icons into the HyperCard system.

2) Selection from a textual menu.
Here objects are selected from within a hierarchical textual menu, again more detail about these objects can be obtained by selection or depiction of object attribute values.

3) Selection can also be based purely upon the object attributes of colour, relative size, orientation, absolute position, length, width and picture number. A user specifying the values of one or more of these attributes either by depiction or selection can then locate pictures based upon their numeric or textual values.

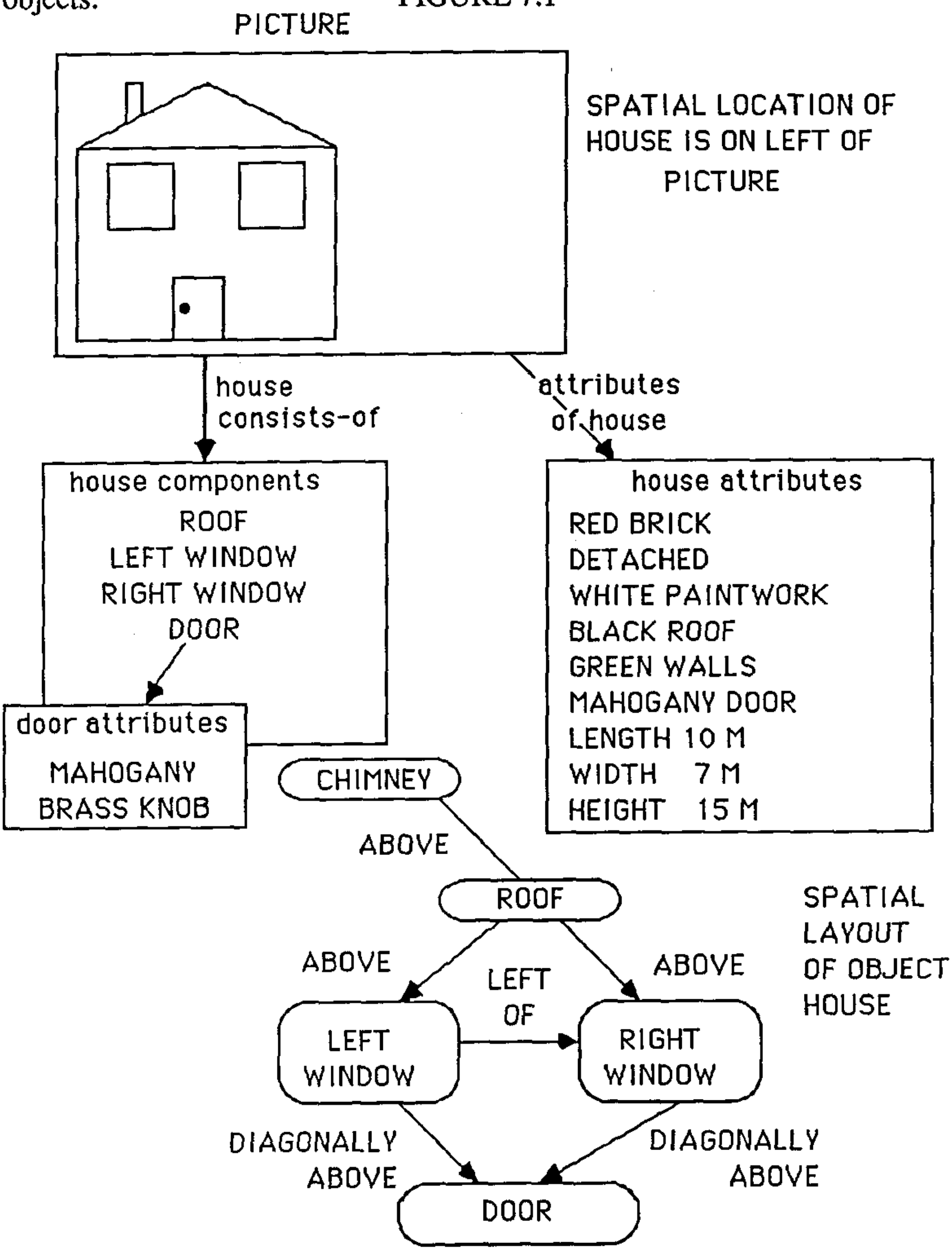
4) Selection of a picture by visually browsing through the pictorial database.

There are two forms of user query for which it is suggested that these four selection techniques are particularly suited. Firstly, queries to find all pictures which match a specific requirement (such as all brown hats within a pictorial database), and secondly queries based upon the users memory of the appearance of objects contained within the pictures. When recalling a picture, users remember objects and/or attributes associated with the picture and it is suggested that using a system such as that demonstrated in this chapter, users can then base their search upon this remembered visual information. Provided that the use'sr can select icons matching the object that they require, can depict absolute spatial positions, sizes, orientations, lengths or widths, or can enter numerical values for the attributes of the objects into the relational database provided, then matching information from the pictorial database can be found and pictures containing this information can then be displayed. The purpose behind this database management system is not to extract textual information about objects; rather it is to find pictures which match the user's object attribute queries, whether either specified visually or textually.

7.5 STRUCTURE OF INFORMATION CONTAINED WITHIN THE PROPOSED DATABASE MANAGEMENT SYSTEM

Complex objects (which can be defined as abstract representations of real world objects (Klaus, Neuhold and Schrefl 1989)) can be decomposed into simpler objects.

FIGURE 7.1



OBJECTS THEIR ATTRIBUTES, AND THE DECOMPOSITION OF THE OBJECTS INTO SIMPLER OBJECTS

Both the complex objects and their constituents have visual attributes upon which searches can be made for specific pictorial information. A proposed example of the decomposition of a complex object into its composite objects, attributes and spatial arrangements is shown above.

Here users could indicate that the house they are looking for from a pictorial database has the colour attribute of green. The position of the house could also be visually depicted on an interface. It is suggested that a database management system can be developed which would search for objects (for example houses) with the attributes specified by the user and in the spatial positions depicted. The components of the house in question (such as windows and doors) can also be used in the search and these objects in turn have visual attributes by which specific pictures containing them can be located.

The basic structure of such a system is a semantic network in which the descriptions are placed in a knowledge base. The user searches for an object by specifying the object and/or its visual attributes. A system such as this is flexible in that it allows the user to find pictures based upon the depicted spatial locations of objects in the picture, as well as other visual attributes of the objects providing a powerful combination of search procedures.

7.6 THE HYPERCARD PICTORIAL DATABASE MANAGEMENT SYSTEM

Apple HyperCard was selected as the the most suitable system available for the construction of the pictorial database management system. This was in part due to the ease of construction of icons and the ability to display them to the user. Additional factors were that an Oracle relational database could be attached to the HyperCard system and a source of pictorial information in the form of a laserdisk was available. Speech was also attached to the database management system using a Macrecorder. The first step in the construction of the database management system was the manual encoding of the visual attributes of the objects within the pictures into the Oracle relational database in accordance with the principles detailed in the last chapter. This stored textual information could then be later directly compared with the user's depicted or textual query.

Having textually encoded attributes of the objects into the relational database, the searching techniques for pictures within the pictorial database were divided into four basic subdivisions, visual depiction of information, textual selection of information, selection of textual information from the relational database and direct physical control of the laserdisk using an on-screen laserdisk controller. The Laserdisk used as the picture source was the PAL formatted Vincent Van Gogh laserdisk containing the life and works of the Painter narrated by Leonard Nimoy.

7.6.1 The Nature of the Pictures Present Within the Laserdisk

Pictures were selected from the laserdisk showing many of the paintings of Vincent Van Gogh. Visual searches for these pictures could be made based upon both representations of physical objects and the depiction of the values of object attributes. Objects contained within the pictures were however of varying size and resolution such that physically small objects often occupied a large proportion of the picture irrespective of their actual real-world physical dimensions, therefore relative size rather than absolute size was chosen as a visual attribute. This attribute of relative size was based upon the size of the screen displaying the information rather than the actual physical size of the object concerned. For example within one frame of the laserdisk a painting of a hat occupied the entire picture, and this was then classified as having a large relative size irrespective of its actual real-world size. Thus using the relative size attribute, a hat had the same relative size as a building provided that it occupied the same physical area on the screen. In using relative size as an attribute by which to select pictures, users therefore needed to be aware of the approximate physical area occupied on the screen by the object required. Absolute rather than relative spatial analysis was also used to identifying specific pictures, based upon object's absolute positions in a picture rather than their relative spatial arrangements with other objects. This was in accordance with the findings in chapter 5.

The attribute values of objects encoded in the Vincent Van Gogh pictorial database were:-

1) the name of the Object
2) its Relative Size
3) its Colour (where applicable)
4) its Orientation
5) the Absolute Position of the object on the screen
6) the Length of the Object
7) the Width of the Object
8) which frame of the laserdisk it was contained on (ie which picture)

The shape of object was not included into the database because of the difficulty in matching between the users shape query and the object information stored on the laserdisk.

Within the database management system, objects could be selected based upon their textual names or by visually choosing from a menu of matching object icons. Objects could also be selected from the database by spatially arranging these icons on interactive interfaces and by visually depicting spatial positions, relative sizes, orientations, lengths and widths. However at present colour can not be visually selected since Apple HyperCard displays information only in monochrome, therefore the colour attribute of an object could only be specified in an object attribute query by choosing it textually from within the Oracle relational database interface.

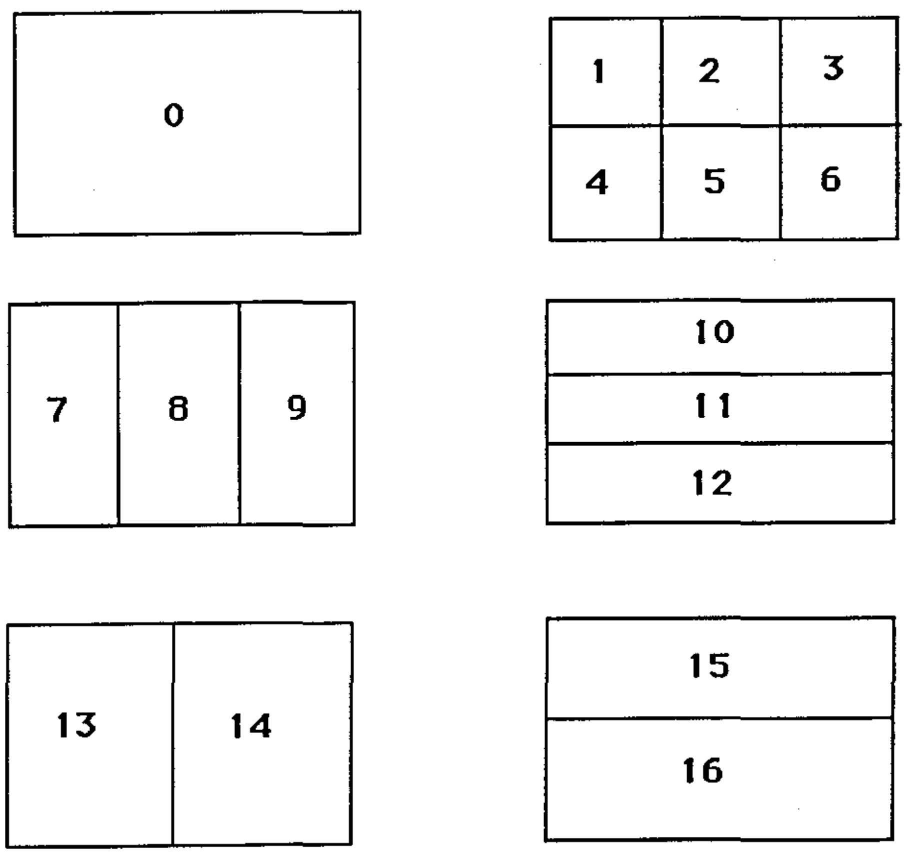
One of the major reasons for using textual information to represent the values of object attributes was the ease with which such information could be stored in a relational database. Having depicted or selected object attribute values, the database management system then compares these values with information contained in the relational database. Detailed use of all but the menus for the spatial selection of information are contained in appendix 4.

7.6.2 The Spatial Menus

Within the HyperCard pictorial database management system there are four methods by which users can specify the absolute spatial location of an object that they require from the laserdisk. The first is to enter the spatial values of the object being searched for directly into the SQL interface, these values ranging from 0 to 16. This can be achieved either by typing the value into the appropriate field directly or by clicking on the spatial area name field from which the available values are displayed.

The second alternative is the visual selection from the six screen possibilities shown below. Due to the varying sizes of the objects stored on the Van Gogh laserdisk, objects within individual pictures on the disk can occupy either the whole screen, a large proportion of the screen, or even a small section of the total screen area. The screens shown below allow the user to specify a combination of size and absolute spatial location by indicating both the approximate size and the spatial location of the object required from the choice below. For example if they remember (or wish to identify) objects which occupy the whole screen then area 0 should be selected from the choice below.

FIGURE 7.2

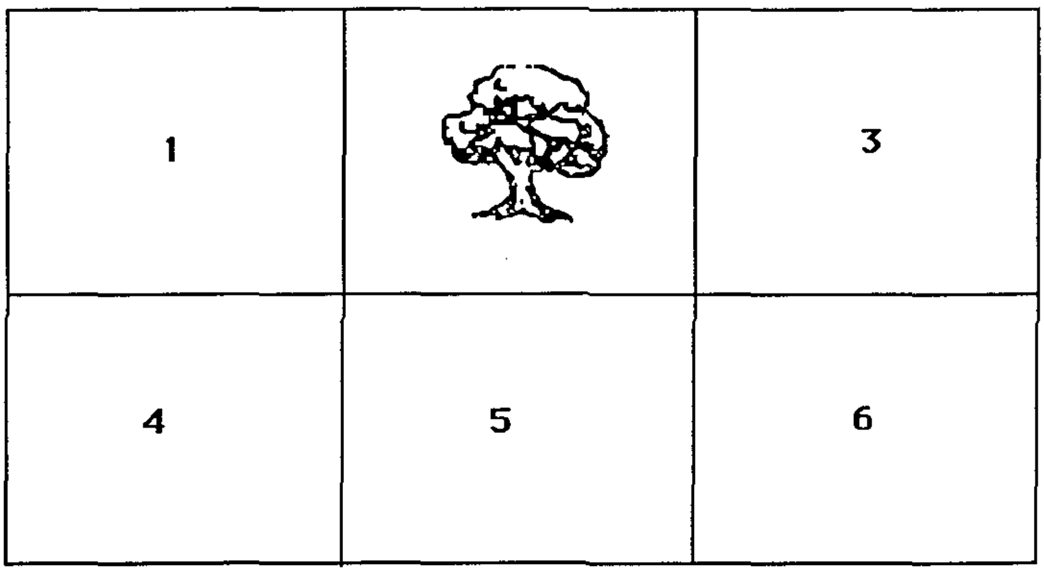


SPATIAL AREA POSSIBILITIES SCALED TO THE SCREEN SIZE

Alternatively if the object required occupies the top third of the screen then area 10 should be selected. Thus simply by clicking on an area, the appropriate numerical value (in the first case 0 and the second case 10) will automatically be placed into the appropriate field of the SQL select statement, so that matching pictures can be displayed upon selection from within the relational database.

A third alternative is the facility to enlarge each of the six screen areas above and paste an icon onto it. This is shown below. Here an icon of a tree has been pasted into area two of the screen. After this the object name 'TREE' is automatically placed into the object name field of the relational database as well as the value of 2 being placed in the spatial area field. This combination of attributes can be used to locate pictures from the laserdisk which match the user's visual recall of a picture that they remember

FIGURE 7.3

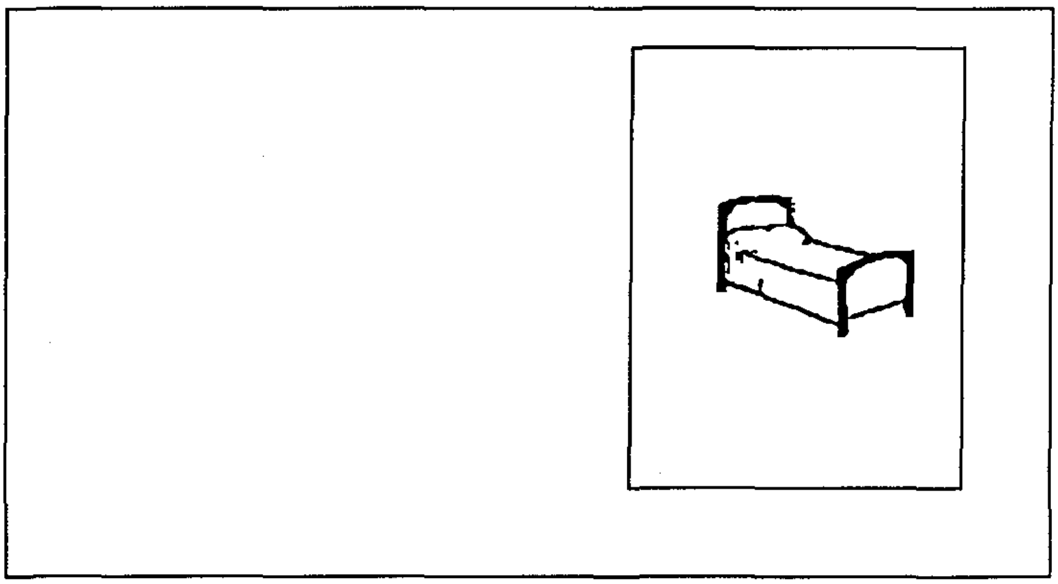


ICON OF A TREE PLACED IN SPATIAL AREA 2

In addition to searches based upon combinations of object name and spatial location, the relational database also allows the user to locate all objects which occupy a specific location by entering the numeric value of an area without the accompanying object name. Therefore entering the value '1' into the spatial area field of the Oracle relational database interface finds all objects within pictures which are located in the top left hand corner of the television screen occupying approximately a sixth of the screen.

The fourth alternative is the ability to specify the absolute spatial location of an object by indicating the spatial area in which the centre of its bounding box lies, this is demonstrated in figure 7.4.

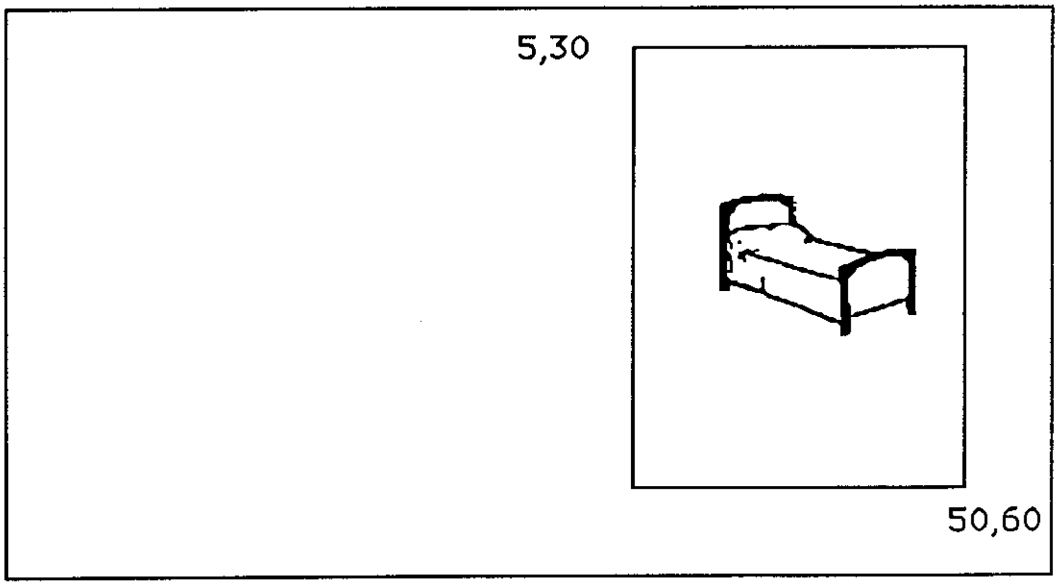
FIGURE 7.4



SPECIFICATION OF THE ABSOLUTE LOCATION OF AN OBJECT (IN THIS CASE A BED) BY DEPICTING THE POSSIBLE RANGE OF ITS BOUNDING BOX CENTRE

In the example shown, a generalised icon of the object bed has been selected and thus the word 'BED' is placed into the name field of the Oracle relational database interface. This is combined with the user's specification of the possible positional range of the bed in the picture required, which is indicated by enclosing it within a rectangle drawn on the interface by the user. Having drawn the rectangle by pointing to its top left and bottom right-hand positions, the coordinates of these positions combined with <= and >= signs are entered into the x and y coordinate fields of the relational database. Thus for example if the coordinates of the rectangle are the following:-

FIGURE 7.5



COORDINATES OF RECTANGLE ENCLOSING BED TO BE FOUND

Then these values are entered into the SPATIAL_X and SPATIAL_Y coordinate position fields of the Oracle relational database interface, from which all beds having positions within these coordinate ranges can be located and displayed to the user. Part of both the Oracle relational database interface and the coordinates from the diagram above are shown below (within HyperCard, coordinate positions are measured from the top left hand corner).

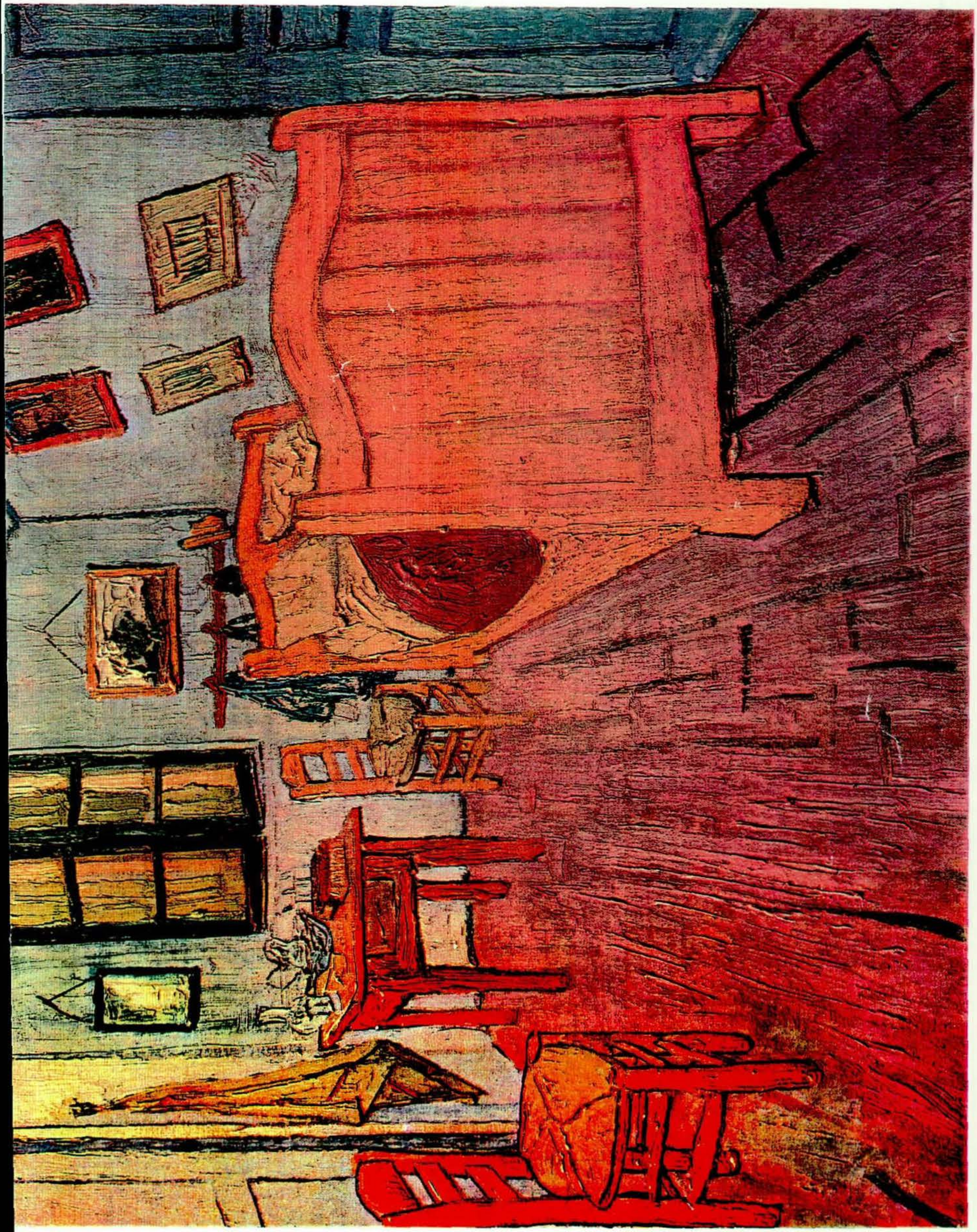
FIGURE 7.6

OBJECT_NAME	BED		
SPATIAL_X	>=5	AND	<=50
SPATIAL_Y	>=30	AND	<=60

PART OF THE ORACLE RELATIONAL DATABASE INTERFACE WITH THE COORDINATE RANGES SHOWN

This technique of object selection and specification of the possible range of the

FIGURE 7.7

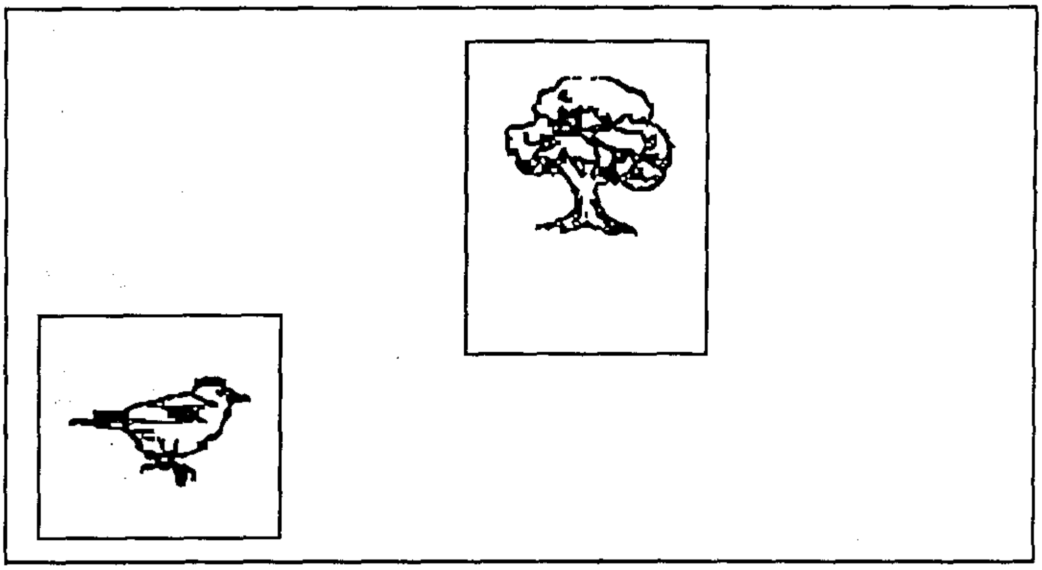


MATCHING VAN GOGH PICTURE TO THE SPATIAL DEPICTION.

7.6.3 Two Objects Within a Picture

One extension to the system currently being developed is to allow two or more objects to be specified within a picture. This can be achieved by allowing the user to specify attributes for two sets of object fields and then having the relational database identify all pictures which match the chosen object combinations. Again, as before, icons can be selected and placed in appropriate spatial positions. For example in the case below a picture is being searched for that has a tree in the top centre part of the picture as well as a bird in the bottom left-hand part of the picture both of whose positions (bounding box centres) are in the ranges shown.

FIGURE 7.8



ICONS OF A TREE AND BIRD PLACED IN SPATIAL RANGES SHOWN

Alternatively the user can also directly enter appropriate textual values for the names and positions into the relational database interface. In this way combinations of objects and their attribute queries can be entered. In the example above a bird icon has been selected with its position within a picture depicted as being in the lower left hand corner of the interface. This is combined with a tree icon depicted in the spatial range shown. This specification of combined object attributes should find matching pictures from the laserdisk containing the two objects in the positions shown.

7.6.4 Practical problems with the Apple HyperCard database management system

Practically the major handicap of the Van Gogh system is the inability of the Oracle interface to specify a combination of queries within a single select statement. Thus at present the Oracle system does not support the ability to place logical combinations of queries into a single field, this therefore restricts the ability to specify a range of values easily into the system. For example the following is not allowed within a field of a single select statement.

FIGURE 7.9

SPATIAL_AREA | >4 .AND. <7 |

ILLEGAL LOGICAL OPERATORS IN ORACLE

This inability to specify ranges of information restricts the user to specifying a single value entry into a field. However allowable single value specifications include those such as

FIGURE 7.10

SPATIAL_AREA | >4 | (spatial area greater than 4)

and

OBJECT_NAME | %F% | (the letter F contained in the object name field)

LEGAL LOGICAL OPERATORS IN ORACLE

These specifications are allowed provided that logical operators OR, AND or NOT are excluded. Therefore although queries such as "Find all flowers greater than size 5" or "all objects starting with the letter F" can be specified, the current implementation of Oracle does not allow queries such as "find all flowers between the size of 5 and 8" or "all objects which begin with the letters F or T" within a single select statement. Thus under these circumstances the user is forced to make a subjective choice as to which attribute value most closely represents the value of the attribute of the object that they are looking for. This often forces the user to make incorrect judgments about the value of the attribute that they require thereby requiring further examination of the system in order to identify objects with slightly different attribute values. Practically this problem has been largely overcome by duplicating identical field names and specifying different ranges within each of the fields.

The second practical problem with the Van Gogh database management system is that when initially textually encoding object descriptions and attribute values into the Oracle relational database certain assumptions are made about the classifications of attribute values by the database management designer. These are that the eventual user of the system will be able to identify pictures based upon the classifications specified. For example the designer may classify an object within the Van Gogh database system as being of size 5 whereas the user may classify such an object as being only of size 3 or 4. Thus there may be differences between the designer's numerical value for an object attribute and the value of an object attribute being searched for by the user. However this problem should be reduced if such attribute values are displayed visually, for example in the case of size, users can more closely visually compare the area occupied on a screen with their memory of the object's size than textually specifying the numerical value representing that area. This problem has also been overcome by the specification of numerical ranges for each attribute value visually depicted.

Another practical problem with the Van Gogh HyperCard system is the inability to "capture" pictures from the laserdisk for use as icons in the pictorial database management system. These icons could then be used to locate objects on the laserdisk. The ability to "capture" pictures from the laserdisk would enable a much closer visual similarity between icons present within the HyperCard iconic menu system and objects present within individual frames on the laserdisk. Currently the only method of entering iconic representations into Apple HyperCard is by using a grey scale scanner to digitise pictures from books or magazines.The ability to capture pictures from the laserdisk could be achieved with a commercially available frame grabber, unfortunately this was not available at the time of writing.

A further problem with the Van Gogh system in its present state is the inability of

Apple HyperCard to manipulate pictorial information and have the numerical degree of this manipulation recorded. Although the positions of icons on the screen can be easily compared with textual data stored within the relational database, a system which allows the user to alter the size, colour and orientation of an icon interactively and extract the resultant information for comparison with a pictorial database is difficult with the current version of HyperCard. For example on choosing an iconic representation of an object there is no ability to easily rotate an icon within HyperCard, thereby more closely visually matching the object present within a picture.

FIGURE 7.11

ICONIC REPRESENTATION OF A FLOWER | ICONIC REPRESENTATION OF A FLOWER ROTATED

Thus at present rather than allowing icons to be manipulated, pictorial information is depicted and these visual depictions are then converted into textual values which are entered into fields of the Oracle relational database.
Finally, since the present implementation of HyperCard does not allow colour or the existence of resizeable screen displays (called cards in HyperCard), this again restricts the ability of the user to select the colour attribute visually or to represent iconic representations in a more lifelike manner.

7.6.5 <u>Possible extensions to the HyperCard database management system</u>
One proposed extension to the Van Gogh pictorial database management system is to enable the user to select pictures by pictorial manipulation of the attributes within a picture. Thus such a system would allow the user to manipulate both icon and the orientation, size or colour of an icon representing the object within a picture. Under these circumstances an icon could be selected, placed in the correct spatial area, increased in size, rotated and its colour changed, in order to more closely visually match an object within the pictorial database. The textual values of these attributes could then be automatically extracted (such as for

example colour red and relative size 4) and compared with values stored within the relational database in order to find matching pictures to the depicted visual query.

Another possible extension to the pictorial database management system is the inclusion of additional objects in a depicted query. This is difficult within the present implementation of HyperCard since the greater the number of objects the larger the number of corresponding attribute fields, so that in the case of queries containing three or more objects the Oracle interface screen becomes cluttered since the interface is fixed in its size. Both the ability to expand the available screen area and to visually manipulate icons are to be provided in the new version of HyperCard (version 2.0).

A longer term proposal is the automatic analysis of pictures within a pictorial database eliminating the need for manual encoding of attributes, the resulting textually abstracted information being stored in the Oracle relational database. Many projects currently exist with the computer studies department at Loughborough University which attempt detailed analysis of pictures and it is hoped that these can soon be connected to the HyperCard system.

7.7 <u>CONCLUSIONS</u>

Within this chapter a practical database management system has been implemented in which various techniques can be used to locate pictures present within a pictorial database. These include visual selection of representations of objects within the pictorial database and depiction of their attribute values. Since users have been found to remember the spatial location of an object within a picture (see chapter 5), and it has been suggested that visual depiction of the values of this attribute on the user-interface allow pictorial information to be specified (see chapter 6), then it is suggested that users can use this combination of techniques to locate pictures from a pictorial database. The HyperCard database management system developed has been designed to take account of these psychological findings and concepts of depiction.

CHAPTER 8

CONCLUSIONS

New techniques of analysing pictorial databases exploit programming methodologies such as rule-based systems and object-oriented representations of pictorial information. In general these methodologies provide extensions which seek to overcome textual relational database descriptive methods of locating entities within pictures. An ability to represent and manipulate pictorial representations of complex objects (which can be seen as an assembly of objects, each containing their own properties and connected by relationships) by using such an object-oriented methodology is desirable in order to identify pictures containing these objects. Important facilities in support of these object-oriented methodologies for describing pictures include the organisation of object representations into generic hierarchies and description of the object attributes within a textual database query system.

Such facilities allows the tuples in a database management system to be represented in a graphical data space. A goal of these object-oriented methodologies is to allow the user to refer to pictorial information within a database via graphical methods such as spatial search, or via symbolic methods such as a visual retrieval statement for the values of object attributes within a picture. To achieve this it is suggested that a mapping can be provided between objects and icons. This mapping can be provided via a link which logically connects the two. When objects and an icon are linked, a selection of the icon implies a selection of the associated objects, these links are created by associating one or more tuples within a relational database with an appropriate icon.

It is suggested that there are two types of associations. The first type, the specific association, links a specific tuple to a specific icon. This is most useful for recording symbolic information about visual data, such as entering a description of the subject matter of a particular photograph. Here the icon is specific to a particular object within a picture in the database. The second type of association, the class association associates a class of tuples with an icon, here selection of an icon selects a class of objects. For example selecting a generalised icon of a flower will select all flowers from a pictorial database rather than a specific type of flower from within an individual picture.

These systems can be expressed in a graphical form so that the information is more accessible and its structure more obvious than in conventional textual database systems. In this way users can find the information they seek without having to specify it precisely, or know exactly where in the pictorial database it is stored. Users can therefore retrieve data from these systems by manipulating pictorial representations of objects on the user interface. This approach permits many types of questions to be answered without the need for a keyboard, although a conventional textual query language can also be provided.

However, a limitation of such systems is the required involvement of the database administrator in setting up the iconic representation of the object data. While the level of effort and expertise required is no more than that expended in setting up the data definition language of a conventional database management system, it would be desirable where possible to have this activity performed in a more automated manner using expert systems and pattern matching techniques to analyse and abstract information from pictures.

An important part of any visual selection technique is the spatial metaphor. In general the spatial mode of thought enables subjects to connect different aspects of patterns into a coherent whole. The logical mode permits the construction of sequences, detection of inconsistencies and the extraction of correct interpretations. The logical mode is better than the spatial in analysing the details of complex situations, whereas the spatial mode is better at taking account of visual information in constructing the global picture. Spatial data is that whose relationships depend on their location, rather than their function.

This thesis has attempted to show that the spatial metaphor is important in recalling pictures and that spatial search techniques can be used to search a pictorial database. However in contrast to earlier search techniques, a visual rather than textual search method has been suggested in which users indicate the physical attributes and values (whether numerical or textual) of such attributes on a user-interface. The database management system then attempts to match these visual user specified queries to textual abstractions of information stored within its database. This technique allows users to make use of their visual memory of the form of objects within a picture rather than being restricted to complex textual statements describing them. One of the most important of these attributes is the spatial information inherent within a picture. Psychological evidence has shown that users remember objects spatial positions and many of the other physical attributes of objects within a picture without conscious attention having been paid to them.

It has been postulated that visual spatial techniques of searching pictorial databases can be subdivided into two basic forms, absolute and relative. Absolute search techniques rely upon identification of the absolute positions of objects within a picture. For example under these circumstances users may remember that the picture they wish to locate has an object they remember located in the top right hand corner of the picture. In specifying the object representation by selection either from a textual or iconic menu, along with its approximate spatial location within a picture, should thereby enable users to find all pictures from the pictorial database which match their specified spatial queries.

Relative forms of search rely upon users being able to identify the relative spatial arrangements between objects within a picture. This relative spatial arrangement of objects may be positioned anywhere within a picture. Here it is suggested that users can indicate on an user-interface how the objects are spatially oriented to each other rather than detailing their absolute position within the picture frame. Having specified this information, a database management system then compares these user queries with textual information stored in a textual database which has been previously abstracted from the pictures stored in the pictorial database.

Using a combination of these two techniques, users are able to identify pictures which match their visual memory of the appearance, the spatial locations and the spatial arrangements of objects within pictures. It is suggested that this information in addition to other object's visual attributes attributes and their corresponding values should assist users in locating pictures from within a pictorial database. Thus visual attribute information is depicted by the user on the user interface and corresponding pictures containing the matching pictorial information can be retrieved from the pictorial database by the database management system and then displayed to the user. Work within the previous chapter has shown that such a system can be constructed and that pictures (in the form of frames within a laserdisk) can be located by the user depicting information onto a macintosh screen representing the user-interface.

CHAPTER 9

REFERENCES

Abel, D.J. (1988). Relational data management facilities for spatial information systems. Proceedings of the Third International Symposium on Spatial Data Handling, Sydney, Australia. 9-18.

Ahuja, N. (1983). On approaches to polygonal decomposition for hierarchical image decomposition. Computer Vision, Graphics and Image Processing, 24, 200-214.

Allport, A. (1989). Visual attention.. In M.I.Posner(Ed.), Foundations of Cognitive Science (pp 631-682) MIT Press.

Antes, J.R. (1978). The time course of picture viewing. Journal of Experimental Psychological Review, 85, 249-277.

Assmann, K., Venema, R., and Hohne, K.H. (1986). The ISQL language, a software tool for the development of pictorial information systems in medicine. In Shi-Kuo Chang, Tadao Ichikawa and Panos A. Ligomenides(Eds), Visual Languages (pp 261-284). Plenum Press.

Atwood, G. (1971). An experimental study of visual imagination and memory Cognitive Psychology, 2, 290-297.

Barker, P.G., Najah, M., and Manji, K.A. (1987). Pictorial communication with computers. International Journal of Man-Machine Studies. 27, 315-336.

Barker, P.G., and Manji, K.A. (1989). Pictorial dialogue methods. International Journal of Man-Machine Studies. 31, 323-349.

Batley, S. (1988). Videodisk browsing and retrieval of video resources. Fifth Int. Conf on Technology and Education. 1. 117-120.

Baum, D.R., and Jonides, J. (1977). Cognitive maps, comparative judgments of imagined vs perceived distance. Psychonomic Society Washington D.C.

Belhage, B.(1989). Improvement of spatial understanding in art. Educational Computing (Netherlands). 5. 1-2. 29-34.

Bennett, John. L. (1971). Spatial concepts as an organising principle for interactive bibliographic search. In D. E. Walker (Ed.), Interactive bibliographic search, the user/computer interface (pp 67-82): AFIPS Press.

Bentley, J.L. (1975). Multidimensional binary search trees used for associative searching, Comm ACM, 18, 9, 509-517.

Berlyne, D.E. (1958). The influence of complexity and novelty in visual figures on orienting responses. Journal of Experimental Psychology, 55, 289-296.

Biederman, I. (1981). On the semantics of a glance at a scene . In M. Kubovy and

J. R Pomerantz (Eds) , Perceptual Organisation, Hillsdale, N.J: Lawrence Erlbaum Associates.

Binford, T. O. (1971). Visual perception by computer. IEEE Conference on Systems and Control, Miami.

Board, C. (1967). Maps as models. P.Haggett (Ed), Models in Geography. Methuen & Co ltd. London. 671-725.

Booth, P.E., and South, M. L. (1982). Information Filing and Finding. ELM publications. 20-32.

Bolt, R. A. (1984). The Human interface, where people and computers meet. Massachusetts Institute of Technology. Lifetime learning publications. 8-21.

Bruce, V., and Green. P. (1985). Visual perception, Physiology, Psychology and Ecology :Erlbaum. 3.

Byrne, R. W. (1979). Memory for urban geography. Quarterly Journal of Experimental Psychology, 31,147-154.

Burgess, C. L., Leigh, W., and Ali, D. (1986). A graphical database interface. Proceeding of the 8th Annual Conference on Computers and Industrial Engineering, 11, 1-4, 355-359.

Bylinsky, Gene. (April 24 1989). Managing with electronic maps. Fortune, 111-120.

Calloway, L. J. (3-6th Jan 1989). Incidental learning of spatially organised DSS Interfaces, a case analysis. In Proc of the 22nd Annual Hawaii International Conference on System Science, III, 413-416.

Canavesia, F., and Marion, R. (August 1989). Optimisation of a human interface to access image data bases. CSELT Technical Reports,Italy. 17, 4, 259-263.

Chandor, A. (1977). The Penguin Dictionary of Computers. Penguin Books.

Chang, N.S., and Fu, K.S. (May 1980). An integrated image analysis and image database danagement system. Report TR-EE-80-20, Purdue University ,W Lafayette Indiana.

Chang, N.S., and Fu, K. S. (November 1981). Picture query languages for pictorial data-base systems. Computer. 23-33.

Chang, Shi-Kuo. (November 1981). Pictorial information systems. Computer. 10-11.

Chang, Shi-Kuo. (January 1987). Visual languages: A Tutorial and Survey. IEEE Software. 29-39.

Chang, Shi-Kuo., and Kunii, T.L. (Nov 1981). Pictorial data-base systems. Computer.13-21.

Chang, Shi-Kuo,. Lin, B.S., and Walser, R. (1979). A generalised zooming technique for pictorial database dystems. AFIPS Conf Proc NCC, 48, 147-156.

Chang, Shi-Kuo,. Reuss, J., and McCormick, B.H. (1977). An integrated relational database system for pictures. Proc IEEE Workshop picture data description and management, 49-60.

Chang, Shi-Kuo,. Reuss, J., and McCormick, B.H. (1978). Design considerations of a pictorial database system. Policy Analysis and Information Systems, 1, 2, 49-70.

Chang, Shi-Kuo,. Yan, C. W., Dimitroff, D. C., and Arndt, T. (May 1988). An intelligent image database system. IEEE Transactions on Software Engineering, 14, 5 , 681-688.

Charles, S.J., and Scrivener, S.A.R.S. (1990). Using depictive queries to search pictorial databases. INTERACT 90, The Third IFIP Conference on Human-Computer Interaction. Cambridge England.

Chock, M., Cardenas, A., and Klinger, A.(Nov 1981). Manipulating data structures in pictorial information systems. Computer. 43-50.

Clarisse, O., and Chang, Shi-Kuo. (1986). Vicon: a visual icon manager. In Shi-Kuo Chang, Tadao Ichikawa and Panos A. Ligomenides(Eds), Visual Languages (pp 152-190). Plenum Press.

Clark, A N. (1985). Longman Dictionary of Geography, Human and Physical. London, Longman, 588.

Clark, K.C. (1985). Geographic information systems: definitions and prospects. Special Libraries Association: Geography Map Division Bulletin, 142, 12.

Cod, E.F. (June 1970). A relational model of data for large shared data banks. Communications of the ACM, 13, 6, 377-389.

Cooper, R.M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. Cognitive Psychology, 6, 84-107.

Cooper, L.A., and Shepard R.N. (1973). The time required to prepare for a rotated stimulus . Memory and Cognition, 1, 246-250.

Cowen, D.J. (November 1988). GIS vs CAD vs DBMS, what are the differences?. Photogramm. Eng. Remote. Sens (USA), 54, 11, 1551-1555.

Date, C. (1977). An introduction to database systems. Reading, Massachusetts. Addison-Wesley.

Deregowski, J. (May 1974). Pictorial perception and culture. In readings from Scientific American. Image Object and Illusion. W.H. Freeman and Company, 79-85.

Derogowski, J. (1980). Image object and illusion. Academic Press, 79.

Donelson, W. (August 1978). Spatial management of information. Computer Graphics, 12 ,3, 203-209.

Economopoulos, P., and Lochovsky, F.H.(1983). A system for managing image data. Information Processing 83. Mason, R.E.A.(Ed.), Elsevier Science (North Holland), 89-94.

Egenhofer, M. J., and Frank, A.U. (1988). Designing object-oriented query languages for GIS. Human Interface Aspects. Proceedings of the Third International Symposium on Spatial Data Handling, August 17-19, 79-96.

Evans, G.W., and Pezdek, K. (1980). Cognitive mapping: Knowledge of real-world distance and location information. Journal of Experimental Psychology, Human Learning and Memory, 6, 13-24.

Exler, R.D. (1988). Future trends in multifold spatial database management. Proceedings NCGA Mapping and Geographical Information Systems 88, 45-54.

Fantz, R.L. (1961). The origin of form perception. Scientific American. 204, 5, 66-72.

Fields, C. and Negroponte, N. (October 1977). Using new clues to find data. Very Large Databases. Proc on the 3rd International Conference on Very Large Databases, Tokyo, 156-158.

Frasson C and Erradi M. (May 27-31 1985). Graphics interaction in databases. Proceedings Graphics Interface. 75-81

Gerstendorfer, M & Rohr, G. (1987). Which task in which representation on what kind of interface? In Human-Computer Interaction - INTERACT 87. Bullinger, H.J & Shackel, B.(Eds.), North Holland: Amersterdam.

Gittins, D. (1986). Icon-based human-computer interaction. International Journal of Man-Machine Studies, 24, 519-543.

Glinert, E.P. and Tanimoto, S. L. (Nov 1984). PICT: An interactive graphical programming environment. IEEE Computer, 17, 11, 7-25.

Goodman, N. (1968). Languages of Art. Bobbs-Merill, Indianapolis.

Gould, M.D., Tatham, J.A., and Savitsky, B. (30th November- 2nd December 1988). Applying spatial search techniques to chemical emergency management.. Third International Conference of GIS/LIS Proceedings. Exhibits and Workshops

San Antonia, Texas, USA, 2, 843-851.

Graham, J. M. (1978). An intelligent spatial database system for interaction with a real-time piloting system. National Aerospace and Electronics Conference NAECON 1987, 4, 1375-1376.

Guastello, S J., Traut, M., and Korienek, Gene.(1989). Verbal versus pictorial representations of objects in a human-computer interface. International Journal of Man-Machine Studies, 31, 99-120.

Hanson, A.J., Pentland A.P., and Quam L.H. (Febuary 1987). Design of prototype interactive cartographic and analysis environment. Image understanding workshop proceedings, Los Angeles, 2, 475-82.

Haralick, R. M., and Simon, J.C. (1980). Issues in Digital Image Processing. Publishers Sijthoff and Noordhoff.

Harrison, H. P. (1981). Picture Librarianship. Publishers Library Association Publishing Limited. Page 143-146.

Hasher, L., and Zacks, R. (1979). Automatic and effortful processes in memory. Journal of Experimental Psychology: General, 108, 3, 356-388.

Herot, C F. (December 1980). Spatial management of data. ACM Transactions on Database Systems, 5, 4, 493-514.

Herot, C. F. (May 26-28, 1982). Graphical user interfaces. Human Factors and Interactive Computer Systems. In Y. Vassiliou (Eds) Proceedings of the NYU Symposium on User Interfaces, New York, Publishers Ablex Publishing Corporation, 88-103

Herot, C. F. (1986). Graphical Input through Machine Recognition of Sketches. Computer Graphics, 10, 2, 97-102.

Hinton, G.E. (1979). Some demonstrations of the effect of structural descriptions in mental imagery. Cognitive Science, 3, 231-250.

Hirakawa, M., Monden, N., Yoshimoto, I.,Tanaka, M., and Ichikawa, T. (1986). Hi-Visual : A language supporting visual interaction in programming. In Shi-Kuo Chang, Tadao Ichikawa and Panos A. Ligomenides(Eds), Visual Languages (pp 233-260). Plenum Press.

Hoffman, D.D., and Richards, M. (1984). Parts of recognition. Cognition, 18, 65-96.

Howe, D. R. (1983). Data analysis for data base design. Publisher Edward Arnold, 3.

IBM Technical Disclosure Bulletin. (September 1986). 29, 4.

Jackson, G. A. (1988). Relational database design with microcomputer applications Publisher Prentice-Hall, 1-5.

Jolicoeur, P., Gluck, M., and Kosslyn, S.M. (1984). Pictures and names : making the connection. Cognitive Psychology, 16, 243-275.

Jones,W. P., and Dumais, S. T. (January 1986). The spatial metaphor for user interfaces: Experimental tests of reference by location verses name. ACM Transactions on Office Information Systems, 4, 1, 42-63.

Joseph, T., and Cardenas, Alfonso, F.(May 1988). PICQUERY: A high level query language for pictorial database management. IEEE Transactions on Software Engineering, 14, 5, 630-638 .

Julston, B. A., and Baron, R. J. (1985). A model of mental imagery .International Journal of Man-Machine Studies, 23, 313-334.

Kahneman, J. (1973). Attention and effort. Englewood Cliffs N.J. Publisher Prentice-Hall, 64-65.

Kemper, A., and Wallrath, M.(1987). An analysis of geometric modelling in database systems, ACM Computing Surveys, 19, 47-90.

Kindborg, M., and Kollerbaur, A. (1987). Visual languages and human computer interaction. Proceedings of the British Computer Society HCL 87 Conference, University of Exeter, 175-187.

Klaus, W., Neuhold, E.J., and Schrefl, M. (1989). Visual databases need data models for multimedia data. In Kunii, T.L.(Ed), Visual Database Systems. Elsevier Science Publishers, 433-462.

Kobayashi, I. (1980). Cartographic databases. In Ski-Kuo Chang.(Ed), Pictorial Information Systems Publisher Springer-Verlag, 322-350.

Koffka, K. (1935). The principles of gestalt psychology. New York: Harcourt, Brace.

Kohler ,I. (1964).The formation and transformation of the perceptual world. Psychological Issues, 3, 1-173.

Kohler, H. (1987). The space-concept and the control of space. Human-Computer Interaction INTERACT 87. Bullringer, H.J., and Shackel, B.(Eds), Publishers B.V (North-Holland), 223-227.

Korfhage, R. R., and Korfhage, M.A. (1986). Criteria for iconic languages. In Shi-Kuo Chang, Tadao Ichikawa and Panos A. Ligomenides(Eds), Visual Languages (pp 207-231). Plenum Press.

Kosslyn, S (1987). Cognition in Action. Smyth, Morris,Levy and Ellis. Lawrence Erlbaum. 255-259.

Kosslyn, S. M., and Chabris C F. (1990). Naming pictures. Journal of Visual Languages and Computing. Publishers Academic Press, 1, 77-99.

Kramlich, D.(1984). Spatial data management on the carl vinson. Database Engineering, 7, 3,144-153.

Lansdale, M..W. (1988). On the memorability of icons in an information retrieval task. Behaviour and Information Technology, 7, 2, 131-151.

Larsgaard, M. (1978). Map Librarianship. Publishers Libraries Unlimited INC, 55-89.

Lin, B.S., and Chang, S. K. (1979). Picture algebra for interface with pictorial database systems. Proceedings of Compsac 79, 525-530.

Lien, Y.E., and Harris, S.K. (1980) . Structured Implementation of an Image Query Language.In S.K Chang and K.S Fu. (Eds.), Pictorial Information Systems. Publisher Springer-Verlag, 416-430.

Lippman, A. (1980). Movie-Maps, An application of the optical videodisk to computer graphics. Computer Graphics. 14, 3, 32-42.

Mandler, J.M and Johnson, N.S. (1976). Some of the thousand words a picture is worth. Journal of Experimental Psychology, Human Learning and Memory, 2, 38-48.

Mandler, J.M., and Parker E. R. (1976). Memory for descriptive and spatial information in complex pictures. Journal of Experimental Psychology: Human Learning and Memory, 2, 1, 38-48

Mandler, J.M., Seegmiller, D., and Day, J. (1977). On the coding of apatial information. Memory and Cognition, 5, 1, 10-16.

Marr, D. (1976). Early processing of visual information Philosophical Translations of the Royal Society of London, series B, 275, 483-524.

Marr, D., and Nishihara H. K. (1978). Representation and recognition of the spatial organisation of three-dimensional shapes. Proc R Soc London 200, 187-217 and 269-294.

McCannn, C.A., Taylor, M.M., and Tuori, M.I. (1988). ISIS. The interactive spatial information system. International Journal of Man-Machine Studies, 28, 101-138.

McKusick, V. A., and Clairborne, R. (1973). Medical genetics. Publishers HP Publishing Company.

McNamara, T.P. (1986). Mental representations of spatial relations. Cognitive Psychology, 18, 87-121.

McNamara, T.P., and Hardy, J.K. (1989). Subjective hierarchies in spatial memory. Journal of Experimental Psychology: Learning Memory and Cognition, 15, 2, 211-227.

McNeil, R J. (May 1966). The shell photographic library. ASLIB proceedings, 18, 5,128-37.

Meier, A and Ilg, M.(1986). Consistent operation on a spatial data structure. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8, 4, 532-538.

Meyer-Wegener, K., Lum, V.Y., and Wu, C.T. (1989). Image management in a multimedia database system. In Kunii, T.L (Ed) Visual Database Systems, Publishers Elsevier Science B.V (North Holland), 479-523.

Miller, G.A. (July 1968). Psychology and information. American Documentation, 19, 3, 286-289.

Murch G. (1973). Visual and auditory perception. Publishers The Bobbs-Merrill Company Inc, 1-2.

Naveh-Benjamin, M. (1987). Coding of spatial location information: An automatic process? Journal of Experimental Psychology Learning Memory and Cognition, 13, 4, 595-605.

Naveh-Benjamin, M. (1988). Recognition memory of spatial location information. Another failure to support automaticity. Memory and Cognition, 16, 5, 437-445

Neisser, U. (1968). The processes of Vision. Image, Object and Illusion. Readings from Scientific American. Publishers W.H Freeman and Company. 4-11.

Newman, W., and Sproull, R. (1979). Principles of Interactive Computer Graphics. Publishers McGraw-Hill.

Nordgren, L. E. (December 1986). Managing graphic or spatial data with Filevision. Small Computers in Libraries, 22-28.

Norman, D.A .(1983). Some observations on mental models. In D Gentner and A.L Stevens (Eds.), Mental Models, Hillsdale NJ: Erlbaum.

Norman, K.L., Weldon L.J .,and Schneiderman, B. (1986). Cognitive layout of windows and multiple screens for user interfaces. International Journal of Man-Machine Studies, 25, 229-248.

Noton, D., and Stark, L,.(1971). Eye movements and visual perception.Image, Object and Illusion. Readings from Scientific American. Publishers W.H Freeman and Company, 113-122

Ooi, B. C., Sacks-Davis, R., and McDonell, K. J. (1987). Spatial kd-tree: an indexing mechanism for spatial databases. COMPSAC 87 Computer Software and Applications Conference, 433-438.

Ooi, B. C., Sacks-Davis, R., and McDonell, K. J. (6-10 Febuary 1989). Extending a DBMS for geographical applications. Proceedings 5th International Conference on Data Engineering. IEEE cat no (89 CH 2695-S), 590-597.

Oosterom, P.J.M.(September 12-15 1988). Spatial data structures in geographical information systems. Proceedings NCGA Mapping and Geographic Information Systems. Orlando, Florida, 104-118.

Paivio, A. (1971). Imagery and Verbal Processes. New York: Holt, Rinehart.

Palmer, B. L.,. and Frank A. U.(August 17-19 1988). Spatial languages. Proceedings of the Third International Symposium on Spatial Data Handling. Sydney, Australia, 201-210.

Peuquet, D.J (1984). A conceptual framework and comparison of spatial data models. Cartographica, 21, 4, 66-113.

Peuquet, D.J. (1988). Towards the definition and use of complex spatial relationships. Proceedings of the Third International Symposium on Spatial Data Handling. Sydney, Australia, 211-224.

Pezdek, K.(1978). Recognition memory for related pictures. Memory and

Cognition, 6, 64-69.

Piaget, J. (1975). Die entwicklung des raumlichen denkens beim kinde. Gesammelte Werke 6. Klett Verlag, Stuttgard.

Pizano, A., Klinger, A., and Cardenas, A. (December 1989). Specification of spatial integrity constraints in pictorial databases. Computer, 59-71.

Potter, M.C., and Faulconer, B.A (1975) Time to understand pictures and words. Nature, 253, 437-438.

Potter, M. C., and Levy E.I. (1969). Recognition memory for a rapid sequence of pictures. Journal of Experimental Psychology, 81, 10-15.

Pylyshyn ,Z.W .(1973). What the mind's eye tells the minds brain: a critique of mental imagery. Psychological Bulletin, 80, 1-24.

Robertson, P. K. (1988). Choosing data representations for the effective visualisation of spatial data. Proceedings of the Third International Symposium on Spatial data Handling. Sydney, Australia, 243-252.

Rock, I. The perception of disoriented figures. (January 1984). Image object and illusion. Scientific American. Publishers W.H Freeman and Company, 71.

Rogers, Y.(September 1986). Evaluating the meaningfulness of icon sets to represent command operations, People and computers: designing for usability. Proceedings of Second Conference of the BCS HCI Specialist Group, York, 586-603.

Rohr, G.(December 1984). Understanding visual symbols. Proceedings of the 1984 IEEE Workshop on Visual Languages, Hiroshima, 184-191.

Rohr, G. (1986). Using visual concepts. In Shi-Kuo Chang, Tadao Ichikawa and Panos A. Ligomenides(Eds), Visual Languages (pp 325-348). Plenum Press.

Rohr, G. (1987). When will direct manipulation of databases be an advantage? Proceedings of the SEAS Anniversary Meeting. Edinburgh, 2, 873-881.

Rosch ,E., Mervis, C.B., Gray,D.M., Johnson,D. M., and Boyes-Braem, P. (1976). Basic objects in natural categories. Cognitive Psychology, 8, 282-439.

Samet, H. 1990. (Preface Page).The Design and Analysis of Spatial Data Structures. Publishers Addison Wesley Publishing Company, Inc.

Sanders, R.E., Gonzalez, E.G., Murphy, M.D., Liddle C.L., and Vitina J.R. (1987). Frequency of occurrence and the criteria for automatic processing. Journal of Experimental Psychology: Learning, Memory & Cognition, 13, 241-255.

Schrivener, S., and Schappo, A. (1988). A perceptual approach to picture interpretation. Knowledge Based Systems, 1, 2, 105-113.

Schulman, A.I. (1973). Recognition memory and the recall of spatial information. Memory and Cognition, 256-260.

Shepard, R.N. (1967). K. Spoehr and S.W. Lehmkuhle (Eds.), Visual Information Processing. Publisher Freeman and Company, 173.

Shneiderman, B. (1982). The future of interactive systems and the emergence of direct manipulation. In Book Human Factors and Interactive Computer Systems Proceedings of the NYU Symposium on User Interfaces. Edited by Yannis C. Publishers Ablex publishing, 1-27.

Shneiderman, B. (1986). An Experimental comparison of a mouse and arrow-jump keys for an interactive encyclopaedia, International Journal of Man-Machine Studies, 23, 23-43.

Shneiderman, B. (1989). Designing the user interface: strategies for effective human computer interaction. Addison Wesley Publishing Company.

Simons, W. W .,and Tansey A,. (1970). A slide classification for the organisation and automatic indexing of interdisciplinary collections of slides and pictures. Picture Librarianship Helen P Harrison (Eds.), Publishers The Library Association Publishing Limited, Santa Cruz, University of California,136-139.

Spooner, D. L. (1986). Relational database support for graphic and CAD. Relational Databases State of the Art Report, 14, 5, 211-227.

Srihari, S.N. and Xiang, Z. (1989). Spatial Knowledge Representation. Int. J Pattern Recognition Artificial. Intelligence, Singapore, 3, 1, 67-84.

Sties, M., Sanyal, B., and Leist, K. (1976). Organisation of object data for an image information system. Proceedings of the Third International Joint Conference of Pattern Recognition, 863-869.

Tamura, H., and Yokoya, N. (1984). Image database systems: a survey. Pattern Recognition (GB), 17, 1, 29-43

Tanimoto, S. L., (1976). An Iconic/symbolic data structuring scheme. Pattern Recognition and Artificial Intelligence. Publishers Academic Press, 452-471.

Tansey, L. (January 1975). Classification of research photographs and slides. Library Trends, 20, 3, 417-426.

Tranowski, D. (1988). A knowledge acquisition environment for scene analysis. International Journal of Man-Machine Studies, 29, 197-213.

Underwood, B.J (1969). Attributes of memory. Psychological Review, 76, 559-573.

Vaughan, W. (October 1988). Painting by numbers. Art History and the Digital Image. Birkbeck College University of London.

Veith, R H. (1988). Visual information systems, the power of graphics and video, Publishers Gower Publishing Company , 143-151

Webb, J M., and Kramer, A F.,(1987). Learning hierarchical menu systems: A comparative investigation of analogical and pictorial formats. Proceedings of the Human Factors Society 31st Annual Meeting, 978-982.

Wertheimer, M. (1958), Principles of perceptual organisation . Readings in Perception, In D.C. Beardslee and M. Wertheimer (eds.,) Princeton, N.J. : Van Nostrand.

Wilton, R.N., and File P.E. (1975). Knowledge of spatial relations: A preliminary investigation. Quarterly Journal of Experimental Psychology, 27,(251-257.

Wilton, R.N and File P.E. 1977). The Representation of spatial information in Memory. Quarterly Journal of Experimental Psychology, 29, 383-388.

Winston, P. H. Artificial Intelligence. Publisher Addison-Wesley. Chapter 8, 227.

Wright, R. M. (19810. Arrangement and indexing. Picture Librarianship by Helen P Harrison. Publishers The Library Association Publishing Limited, 131-148.

Yarbus, A.L. (1967). Eye movements and vision. New York Publisher Plenum.

Yoshimura, H., Tanaka, M., and Ichikawa, A., (November 7-9 1983). Visual interface with semantic zooming and panning facilities for geographic information retrieval. IEEE Workshop on Languages for Automation. American Congress hotel, Chicago, 65-71.

Zloof, M. M. (1975). Query by example. IBM. J Watson Research Centre, NCC AFIPS conference proceedings, 44, 431-438.

APPENDIX

APPENDIX 1

INTERACTIVE SYSTEMS AND ICONS

A 1.1 INTRODUCTION

A 1.2 INTERACTIVE INTERFACES

A 1.2.1 Requirements of Direct Manipulation

A 1.2.2 Analysing Databases

A 1.2.3 Cognition within Direct Manipulation

A 1.2.4 Directness within Direct Manipulation

A 1.2.4.1 Semantic Distance

A 1.2.4.2 Articulatory Distance

A 1.2.5 Direct Engagement

A 1.2.6 Representing Direct Manipulation in Databases

A 1.3 ICONS

A 1.3.1 Advantages and Disadvantages of Icons

A 1.3.2 Classification of Icons

A 1.3.3 Iconic Languages

A 1.3.4 Choosing an Icon

A 1.3.5 Icon Based Command Languages

A 1.3.6 System Characteristics

A 1.4 CONCLUSIONS

A1.1 INTRODUCTION

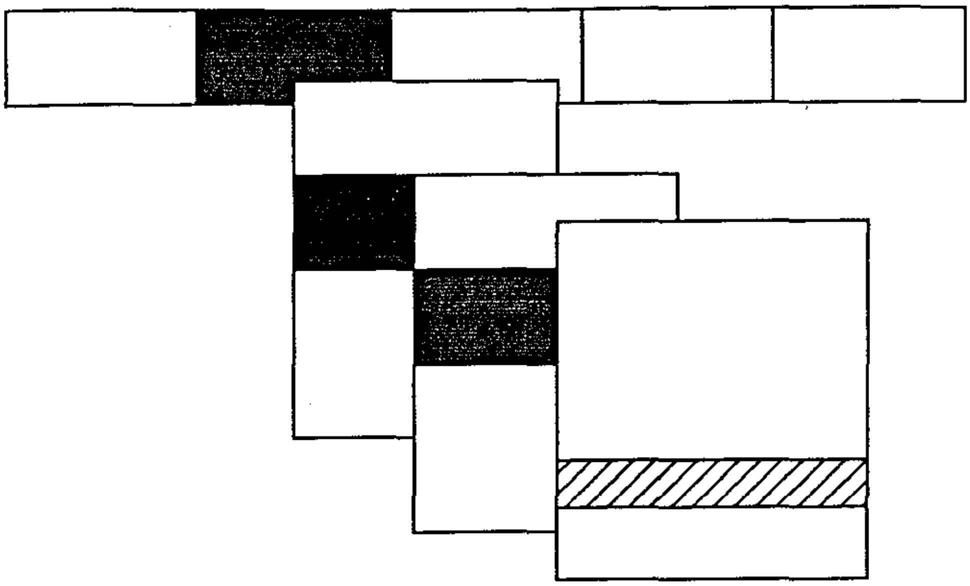
A spatially interactive system requires two basic components, firstly an interface which allows the user to indicate either the spatial location or the spatial relationships between objects from which matching configurations within a pictorial database can be located, and secondly representations (usually icons), which allow the user to specify the object (or objects) to be searched for within the database.

A1.2 INTERACTIVE INTERFACES

Many objects within a graphical user-interface can be referenced by pointing rather than by typing a textual description (Rohr 1987). For example a town may be more readily selected by pointing to it on an interface containing a pictorial map instead of textually entering its name. Selecting by pointing is a major object-oriented feature of graphical presentation because it allows users to select what they see, exploiting both spatial information and visual properties of images. User-interfaces often allow direct manipulation of both properties and attributes of an object within it. This direct manipulation makes these user-interfaces both easy to learn and unintimidating to use.

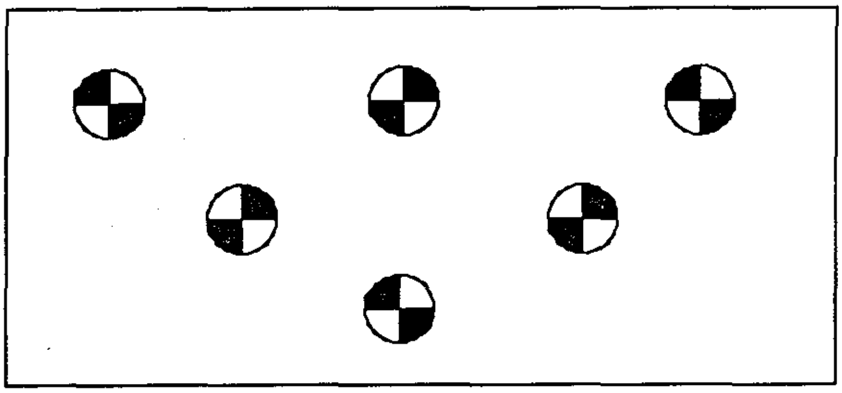
Direct manipulation is always defined as working on graphical objects, however text can be regarded as a graphical object who's graphical properties (font, size, format etc) can be directly and visibly changed. Such operations are defined by pointing at objects arranged in space and then initiating an action usually by some form of clicking action. These objects may then change some of their parameters such as colour, shape or size or alternatively may be moved around the screen. All changes to these objects are immediately visible. The graphical representations used can be abstract and show spatial relations, or they can be more picture like using metaphors such as iconic representations to describe the system component (Rohr 1987). Examples of object attribute manipulation by pointing is shown in figure A1.1.

FIGURE A1.1



SELECTION FROM A PULL DOWN MENU

FIGURE A1.2



SPATIAL MANIPULATION OF ICONS ON A USER INTERFACE

Here in the first case the user is asked to select choices from pull down menu boxes highlighted on the screen. These pull down menu boxes use abstract spatial relations to express menu dependencies. In the second case icons (in this case circles divided into quadrants) are moved around an interactive interface, the positions of these icons representing a particular spatial arrangement of objects within a pictorial database. Both operations are initiated by pointing at the objects and initiating an action by a special clicking event. Such direct manipulation techniques encourage new interaction concepts which are geared towards 'seeing and pointing' rather than 'remembering and typing'.

A1.2.1 Requirements of Direct Manipulation

In order to achieve direct manipulation a high resolution screen is usually required, either with different grey scale levels or preferably in colour. Also needed is some form of pointing device such as a mouse, joystick or touchscreen from which actions can be easily initiated. For the user not to be confused, the semantics of the graphical representation used needs to be carefully mapped to the semantics of the software system thus allowing the relationships between the two to be clearly understood by the users.

A1.2.2 Analysing Databases

Within all forms of database there are three operational levels which have to be analysed. The first level is the selection of parts of a database to be operated on, the second level is the definition of the query and the third level is the operation on the query result (Rohr 1987). The first level the selection of parts of a database can be regarded as a zoom or selected view of the database therefore this has a spatial characteristic and is best suited by a graphical representation. In the second level the structural relations between objects become important and again graphics are the best representation to show both the required attributes and the spatial queries. The third level is mainly determined by the purpose of the result and can be represented either graphically or by a formal table format.

Shneiderman (1982) has suggested that systems using direct manipulation all have the following characteristics in common:-

1) continuous representation of the object of interest.

2) physical actions instead of complex syntax's.

3) rapid incremental reversible operations whose impact on the object of interest is immediately visible.

Using these characteristics it is important to design database management systems which have the following attributes:- the ability for novices to learn basic functions quickly yet having within the system the ability for expert to work extremely rapidly. Such a system should also allow users to retain operational concepts so that similar operations within the system achieve the same results. Users should additionally be allowed to see if their actions are furthering their goals and if not change the direction of their activity.

A1.2.3 Cognition Within Direct Manipulation

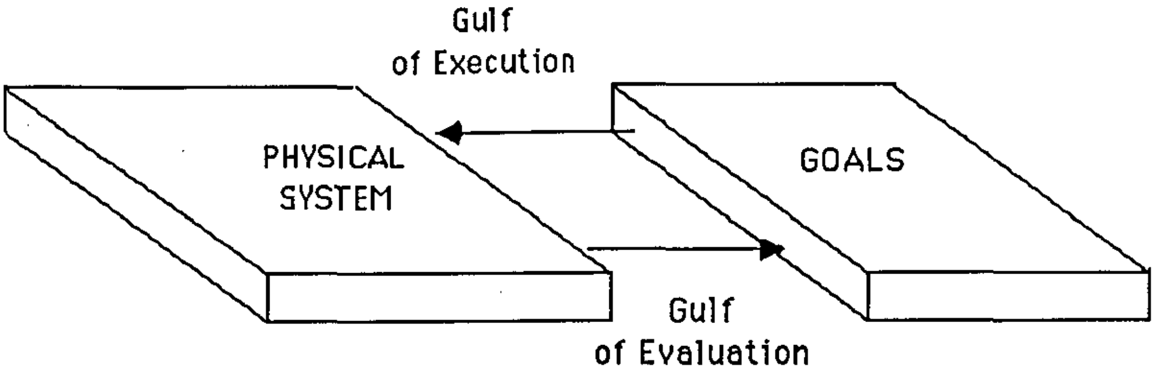
Cognitive models of user behaviour are important in the study of direct manipulation since they enable subjects using these interfaces to do so in a more natural way. Two of the components of these cognitive models are syntactic and semantic knowledge. Syntactic knowledge is the knowledge about details of command syntax, ie what the structure of the commands the user types into the computer actually are. This knowledge depends upon the computer and operating system that the subject uses.

Semantic knowledge can be defined as the principles and concepts that underline the operation of the system the subject is using. Within a word processing package this might include moving a paragraph, deleting a word or centring a heading. Therefore semantic knowledge ranges from low level functions to high level concepts. It is mostly system independent, however the exact nature of the command syntax's available varies. Such semantic knowledge is usually stable in users memory since it is learned through analogy and example and the principles learned can be applied to other software packages and operating systems.

A1.2.4 Directness Within Direct Manipulation

The directness within direct manipulation has two important aspects, distance and engagement (Shneiderman 1989). Direct manipulation distance is the distance between the physical requirements of the system and the ideas behind the system, A short distance means that implementation is easy in terms of translating the ideas into a physical reality. This distance can be expressed in terms of the gulf of execution and evaluation.

FIGURE A1.4



THE GULFS OF EXECUTION AND EVALUATION

These gulfs are minimised by making the commands and mechanisms of the system match the thoughts and goals of the user as much as possible. Schneiderman (1989) suggests that the feeling of directness is inversely proportional to the amount of cognitive effect it takes to manipulate and evaluate a system and that the cognitive effect is a direct result of the gulfs of execution and evaluation. Thus the more the interface minimises the gulf, then the less is the cognitive effect needed and the better is the direct manipulation. The gulf is bridged by two sides, on the system side by the system interface and from the user's side by the mental structure and design. The best direct manipulation interfaces give the feeling that the user is in direct control of the objects concerned, thus rather than linguistically giving commands referring to the objects of interest, representations of the objects themselves are manipulated. Therefore such manipulation of a representation needs to have the same effect and feel as manipulating the actual object being represented. When the user directly engages a system it can be considered that there are two conceptual forms of distance, semantic and articulatory.

A1.2.4.1 <u>Semantic Distance</u>

Semantic distance is the relation of the meaning of an expression in the interface language to what the user wants to say. There are two important points here. Firstly is it possible to say what one wants to say in the language? In other words does the language support what the user wants to do and does it do it in the same way that the user wants? Secondly can the thing of interest be said concisely in a straightforward fashion or must the user of the system explain the problem in terms of complex computer expressions?

The semantic distance can be reduced by either the designer constructing higher-order and specialised languages that move toward making the semantics of the input and output languages match that of the user, or by the user developing competence by building new mental structures to bridge the gulfs.

In many cases the semantic distance can be reduced by making the output show the semantic concepts directly. An example of this is the use of screen text editors rather than line editors where the effects of editing much more closely match the user's semantics. Again the same problem applies, if the system is too specific and specialised then the output displays lack of generality, however if the system contains too much information then the user has trouble learning and selecting among the various possibilities.

The semantic distance is not reduced simply because of automated behaviour. Users often become familiar with a system even if it is difficult to use and they learn to do things automatically once told even if it is not semantically apparent what they are doing. Therefore under these circumstances the deficiencies of the interface are compensated by the constant use of the system. One way in which the semantic distance can be reduced is for users to see the problem in the same terms as the system does and then formulate the problem and the solution in terms of the system rather than in terms of their own ideas.

A1.2.4.2 Articulatory Distance

Articulatory distance is the relationship between the meanings of expressions and their physical form. On the input side the form may be a sequence of key-presses or mouse-clicks from a mouse. On the output side the form might be a string of characters, a change in an iconic shape or a graph diagram.

When using a mouse or joystick to move a cursor on the screen then such an action mimics the desired motion the user wishes the cursor to make, and this action can be regarded as an articulatory direct interaction. Therefore one way to provide articulatory directness on the input side is to provide an interface that permits specification of an action by mimicking it.

Articulatory directness at the output side is similar. If the user is following the changes in some variable, a moving graphical display can provide articulatory directness. A table of numbers although containing the same semantic information is not articulatory direct. Thus the graphical display and the table of numbers might be equal in semantic directness but unequal in articulatory directness. Articulatory directness is highly dependent upon input-output technology. For example the joystick or mouse is a spatio-mimetic device which provides articulatory direct input for tasks that can be represented spatially.

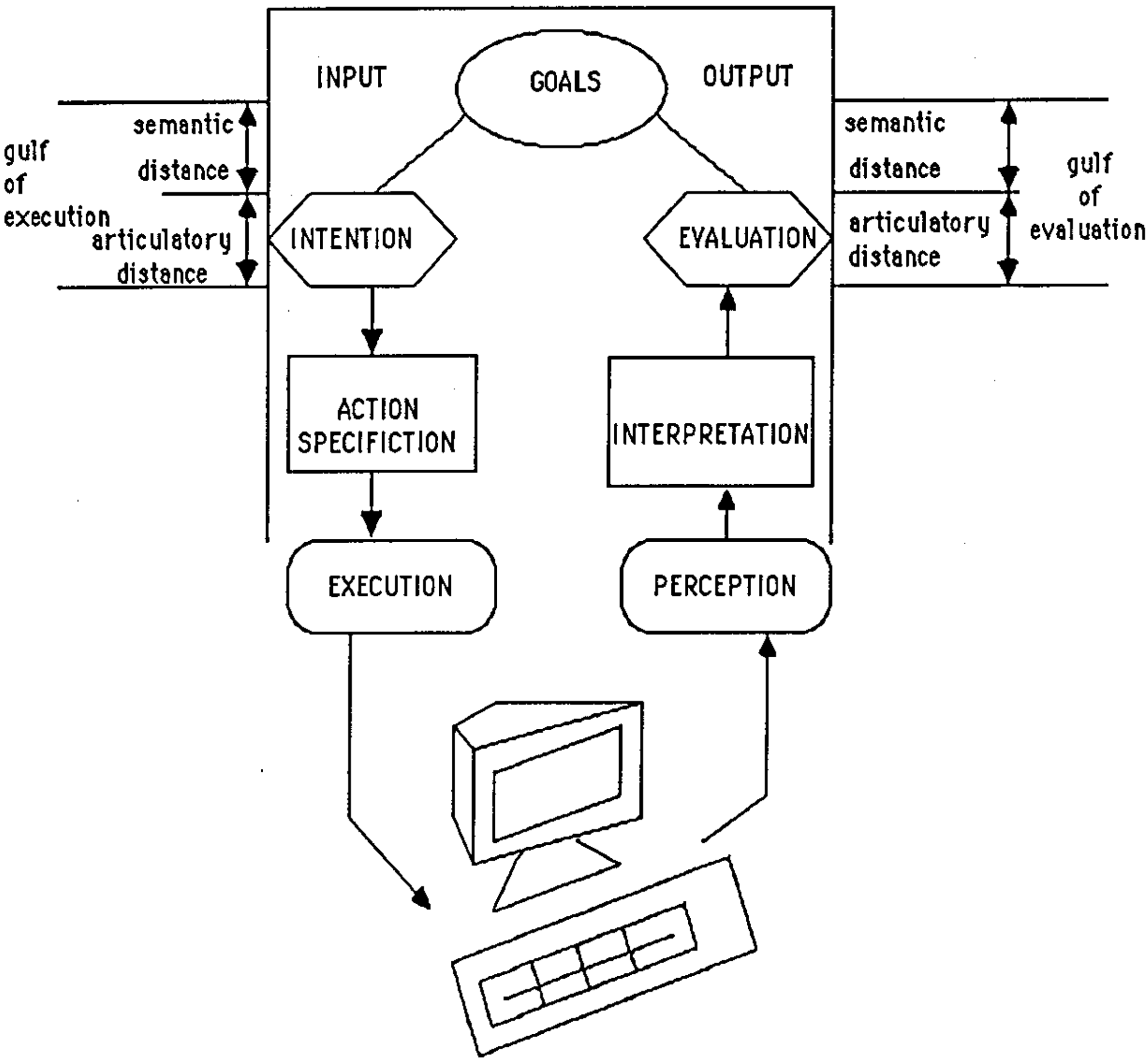
A1.2.5 <u>Direct Engagement</u>

Direct engagement occurs when the user of a system directly interacts with that system, this is particularly important for spatial database applications in which representations of objects are manipulated. Here actions and observations are made directly upon the objects being engaged. Such a direct engagement requires a fast response time so that there are no delays between execution and the results. When using such direct engagements the interface should not interfere with the basic task being undertaken. Therefore for direct engagement the objects on the screen must act and feel like the objects that they actually represent. Whatever changes that occur upon the actual objects must be accurately portrayed on the representations. The use of the same object as both input and output entity is essential to providing objects that behave as if they were the real thing. A rapid response time is particularly important because this both allows for modification of actions as well as the appearance of acting directly upon the objects themselves. Ideally the user should see as little of the system operation as possible

A1.2.6 <u>Representing Direct Manipulation in Databases</u>

Using direct manipulation there are various possibilities for objects relationships and operations which have to be mapped to the user's understanding within a pictorial database management system. One of the most important decisions is whether to use concrete or abstract graphical representations. Concrete symbols show natural relationships between objects and this enables characteristics of the symbols to be found quickly, but here functional dependencies are more difficult to show. With abstract spatial graphics, functional dependencies expressed as spatial relations can be shown, however natural relationships are more difficult to show. Therefore in choosing either concrete or abstract representations the designer of the system has to consider whether natural relationships or functional dependencies are the most important.

FIGURE A1.5



PRACTICAL GULFS BETWEEN EVALUATION AND EXECUTION

A1.3 ICONS

Webster's dictionary defines an icon as "an image; figure; representation or picture" (Chang 1987). These icons often bear a natural resemblance to the objects they represent (Tanimoto 1976). Thus icons are easily learned, remembered and recalled as single units of information which can express information concerning the spatial disposition of objects, their appearance, their texture and their composition. Within the field of computing icons can be defined as predefined flat symbols which can be used to represent objects within a computer interface (Gittins 1986). Therefore such icons can be considered as objects which are placed in two-dimensional space.

The attributes of an icon in question can represent the attributes, the association or the state of the computer system at the time, providing greater information than dialogue within less physical space. The object that the icon represents within the interface can also be either physical or abstract. The key to an icon is a quick identification and understanding of a concept (Clarisse and Chang 1986).

FIGURE A1.6

EXAMPLE OF ICONS

Icons within the computing environment tend to fall into one of two classes, those indicating objects and those indicating processes and processing (Korfhage and Korfhage 1986), however within this thesis icons have been used to indicate objects rather than processes. An object icon represents a physical or logical entity that generally has associated with it a limited set of actions or textual statements. Object icons tend to be easy to understand because of their resemblance to real world objects. Interfaces that contain such objects icons are called object-oriented interfaces.

Such object-oriented interfaces directly benefit users since they are much easier for users to learn and work with than conventional textual interfaces. An object-oriented user-interface is also more appealing because it presents the user with a view of the application that is similar to the way humans deal with the rest of their environment. People handle physical objects such as books, pens and coffee cups or more business related objects such as files or diaries, thus in a non-computerised environment, the user visually finds an object and physically manipulates it so executing its function. Similarly in a computer-based environment iconic representations of objects are manipulated in the same way. For example a book on a screen can be read and its pages turned by mouse clicks on the page corners to either move back or forward through the book. Rohr (1984) has shown that there is a strong relationship between the degree of abstraction of an icon and its type of use. Object icons tend to be strongly pictorial, here the picture of the element is reduced to its simplest form. Some of the characteristics of an object icon include the fact that it usually consists of one component for quick recognition by users and that it is graphical not textual. It also can have only one meaning in the context for which it is to be used. Within a graphic interface users can select one or more icons on a display in order to either invoke processes or manipulate data within their application. Such icons are characterised by type, form and colour, although they can also be classified on the basis of the underlying computer system to which they interface. The main reason that a powerful user-interface is important is that it helps users to understand the structure of the system and allows them to interact with the system effectively. Rohr (1986) found that when both icons and words were used for interactions with a computer system, then subjects using words required considerably more explanation at the beginning of the session than those using icons in order to encode the information into a clearly understandable form.

In the case of the real world icons, then the nature and shape of the icon is used to represent the class of object being searched for. For example an icon of a tree would represent all trees rather than a specific tree contained in a particular picture. (see appendix 4). Here each icon can be used to represent its class (object type) and spatial location, other information about the object the icon represents being provided within a table. Thus the icon's main purpose here is as a marker for the location of a similar class of objects that the user wishes to find within a pictorial database. The actual nature of the icon simply acts as a prompt as an aid to recalling the position of a similar object in the same class within a picture. Therefore a system that uses words as prompts could presumably achieve the same ease of use as a system employing icons. However this is only true where an icon can be replaced by a single word, in other cases, icons can reduce a complex notion to manageable size. Under these circumstances icons not only save screen space but also can be grasped at a glance (Veith 1988). Additionally such icons can in many cases aid the user's recognition of a particular image, particularly in cases where the icon in question closely resembles the object being searched for within a pictorial database.

A1.3.1 Advantages and Disadvantages of Icons

Icons have many advantages over text-based information, these include their ability to indicate to the users something about the nature of their function by their appearance. Thus a novice user may often correctly guess the function of an icon without referring to a manual. Icons use the space provided to design the system more efficiently than text because of their ability to change attributes and still be comprehensible to the user. Icons also improve error reduction by limiting the user to predefined choices rather than selecting the correct text type from memory. Therefore the user is able to determine what functions are allowed.

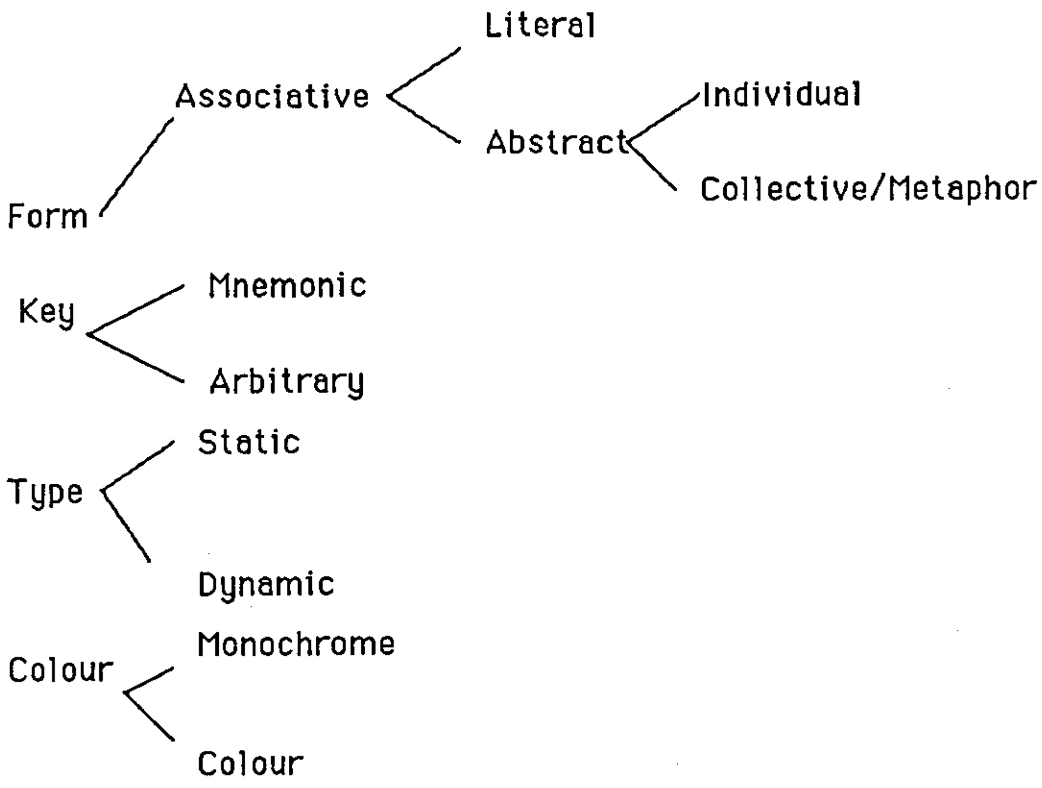
The use of icons allows the designer of the system to use commonly known graphical symbols on the basis of colour and shape and are therefore extremely useful in 'object-oriented' situations where direct manipulation is required. However not all objects within a computer system have an obvious iconic representation. Having an icon to represent certain functions of a program may not inform the user what is happening. Thus when an icon replaces a command there may be parameters that accompany that command and these may be difficult to represent. In order to cater for these parameters the icons may become prohibitively complex or may require textual information alongside in order for it to carry out its function. For example an icon which indicates printing with a pictorial representation of a printer can not easily take account of the font required within the user application. Icons may also indicate to users the wrong characteristics. For example an icon of a printer may indicate a file is being printed to one person, whereas another person may conclude that this means highlight the icon to print the file in question, therefore here there is a difference in users understanding as to what the icon represents.

A1.3.2 Classification of Icons

An icon can either have a characteristic or attributes of the object in question whereby it is called associative, or it can serve as a cognitive key to the subject, whereby the icon is called a key. Both of these serve to identify to the user the nature of the icon (Gittins 1986).

When a collection of icons share a common feature they are collectively known as a metaphor. The alternative is when an icon serves as a cognitive key for an object, is this case it is either "mnemonic" or "arbitrary". A "mnemonic icon" is one in which the icon reminds the user of the nature of the operation. For example concatenation might be represented by the picture of a cat. An "arbitrary icon" does not have any characteristics from which the nature of the object can be guessed.
Chang's taxonomy of icons however classifies them by design and function. By this taxonomy there are three types of icons as illustrated below, these are representational, abstract and arbitrary (Chang 1987). An icon image is chosen to relate to the idea or action either by resemblance (picture), by analogy (symbol), or by being selected from a previously defined and learned group of arbitrary images (sign).

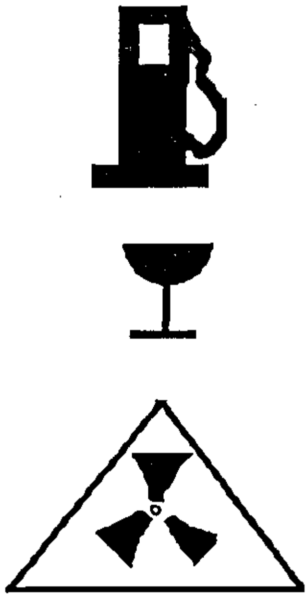
FIGURE A1.7



TYPES OF ICON

FIGURE A1.8

Design	Function
Representational	Picture
Analogy	Symbol
Arbitrary	Sign



A TAXONOMY OF ICONS

A1.3.3 Iconic Languages

Languages in general are constructed in four layers (Korfhage 1986). The innermost layer consists of the set of symbols that form the sentences of the language. In natural languages these are the words of the language. The next layer is the syntax or set of rules whereby the words are put together to form sentences that can be properly parsed. The third linguistic layer, the semantics, provides the framework for interpretation of each sentence, and begins with an attempt at interpretation of each word or symbol and progresses to attempts to interpret sentences and phrases. The final layer is the pragmatics of the language.

Iconic languages must possess these same layers. Firstly there must be an iconography. Secondly there must be a syntax for combining the icons into statements and commands. Thirdly there must be a semantic interpretation of these statements and commands and finally there must be some way of determining whether an iconic command language makes sense in a given environment. In general the two elements that are usually lacking are the syntax for creating statements and commands from icons, and secondly the rules for determining semantics.

A1.3.4 Choosing an Icon

There are several basic steps to choosing an iconic representation. The first aspect is to analyse the use of the icon within the system and try to take account of any future extensions to that use. The second aspect is to design the basic structure of the icon upon which all other icons can be based. Usually other such icons will also be of the same size. Different specific items can then be added to the structure of the basic icon in order to take account of the varying functions.

Icons for information processing have been developed in two basic application areas, business data processing and computer graphics systems. In general business-oriented iconographies have more object icons whereas graphics-oriented iconographies have a better stock of action icons. The iconographies within computer graphics systems are concentrated on the actions that can be applied to objects. The object icons tend to be lines rectangles and circles and there is no way to construct a complex object, name it and assign it a unique icon. Instead complex objects are frequently treated as structures built from the simple graphic elements.

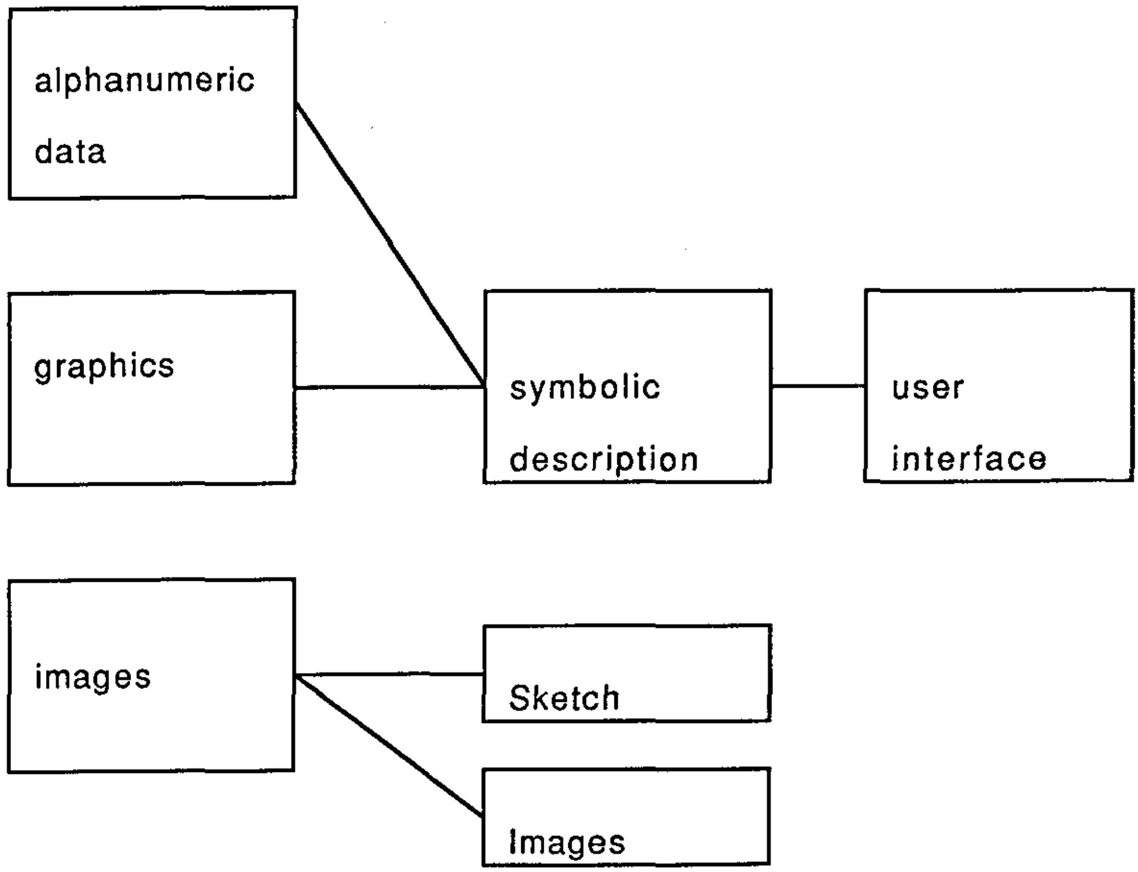
These are then identified dynamically by pointing to them, rather than assigning a named icon to each.

In developing an iconic language for user communication with the system, a wide variety of relationships among the objects, concepts or actions needs to be able to be expressed. Hierarchical relationships are handled in an iconography quite easily, for example opening a file with an icon is a single instruction that immediately exposes the next layer of the hierarchy.

A1.3.5 Icon-Based Command Languages

The advantage of an icon-based command language lies in the fact that a query can be built up in any order the user likes which is important since the perception of the problem and the solution are generally different for different users. The end user's wishes for a direct visual interaction with the computer and they require the ability to query without worrying about the procedural detail (Frasson and Erradi 1985). There are various different categories of database that the users can interact with as shown in figure A1.9

FIGURE A1.9



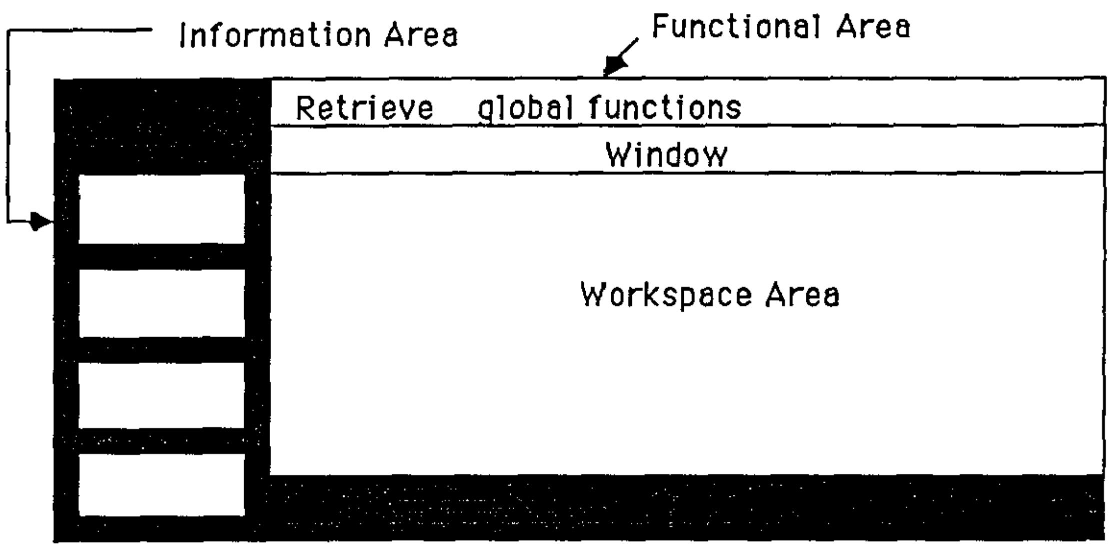
CATEGORIES OF DATABASE THAT THE USER CAN INTERACT WITH

Users who understand their own needs usually wish for a direct interaction with the computer. Here instead of using a tabular form of query formulation, icons can be sketched in order to retrieve relational data. The basic icon-based system consists of two main components, an object interface and an icon-based set of commands. The object interface has a set of objects that the user understands as well as a set of operations that the user can preform on those objects. The actual display area is often split into three areas.

The workspace area is where the icons are manipulated, The information area contains the different properties of the objects, and the functional area gives the different types of operations that can be carried out on the objects within the database. The icons within the application can be selected with the aid of a pointing device.

Frasson and Erradi have applied these principles to medical databases. Here the structure of the database is displayed by selecting an operation within the functional area and the nature of the information dictates what will happen within the interface.

FIGURE A1.10



DISPLAY AREA FOR OBJECT INTERFACE AND ICON COMMANDS

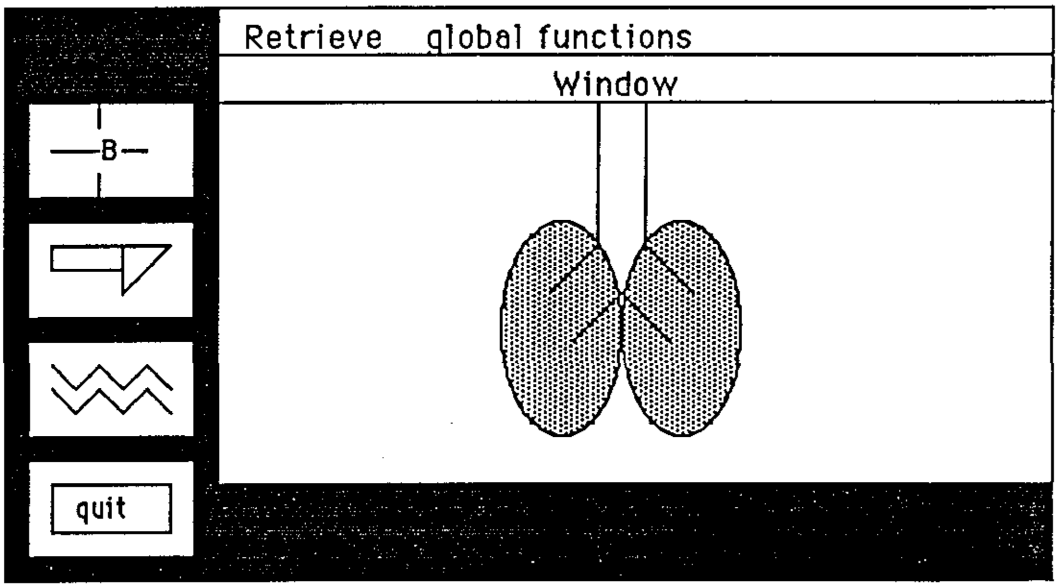
In the case of alphanumeric information all the properties are displayed in the workspace. In the case of an image the set of all the properties of that image are displayed in the functional area. For example in figure A1.11 iconic representations of the lung and the three icons represented on the left of the picture are the diseases of the lung, B standing for Bronchopneumonia, the knife icon represents perforation and the zig-zag lines represent oedema.

The icon-based manipulations that can be carried out for this application include;-

CREATE	Creates an icon for a new object.
MODIFY	Modifies an icon.
DELETE	Erases an icon from the interface.
ZOOM	Enlarges a part of the object.
CUT	Eliminates a property from an object.
HIGHLIGHT	Highlights icons satisfying certain conditions in a composed object.
SHOW STRUCTURE	Gives the iconic representation of the objects in the database and their relationships.

Therefore complicated medical analysis can be executed simply by iconic selection from the choices given.

FIGURE A1.11



AN OBJECT WITH ITS PROPERTIES

A1.3.6 System Characteristics

An iconic interface is a view of all icons available for the user in an application. Icons used for an application can be directly extracted from a set of basic icons or modified using an icon editor. Practically such a system can be represented by figure A1.12A. A practical example of this system uses a medical database. Within this medical database equivalent relational database's relations are:-

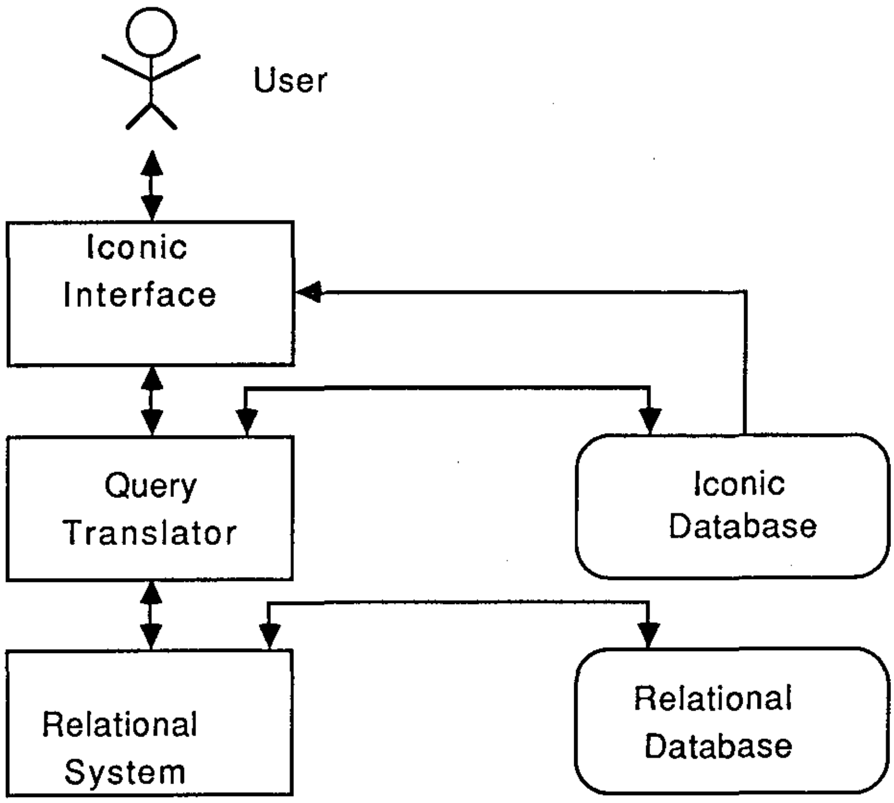
PATIENT (Pat-no, name, address, birth-date)

STOMACH-DIS (Pat-no, dis-no, disease, diagnosis, date, X1, Y1, width)

LUNG-DIS (Pat-no, dis-no, disease, diagnosis, date, X1, Y1, width)

PICTURE (Pat-no, dis-no, frame)

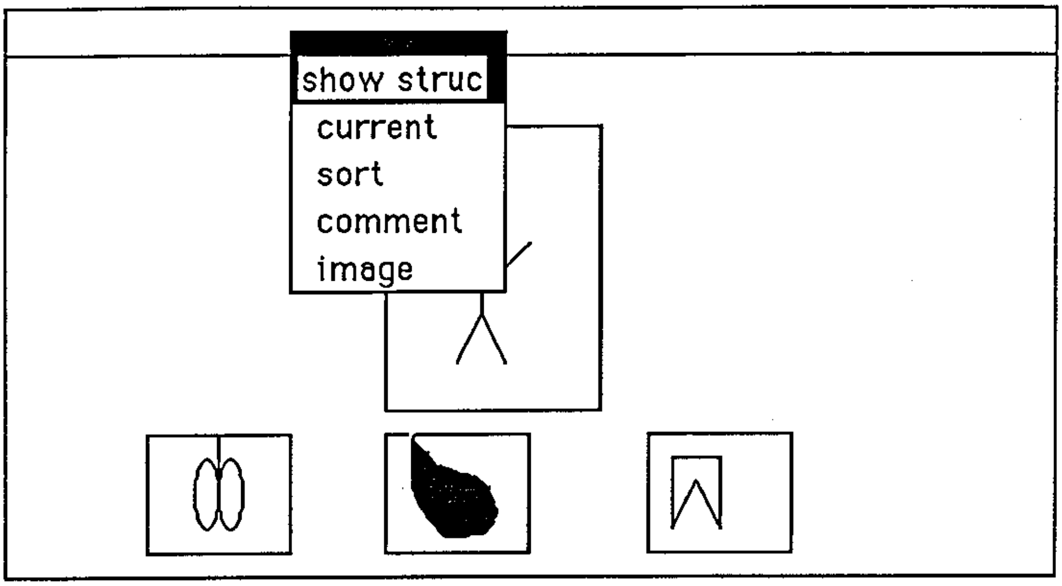
FIGURE A1.12A



ICONIC INTERFACE

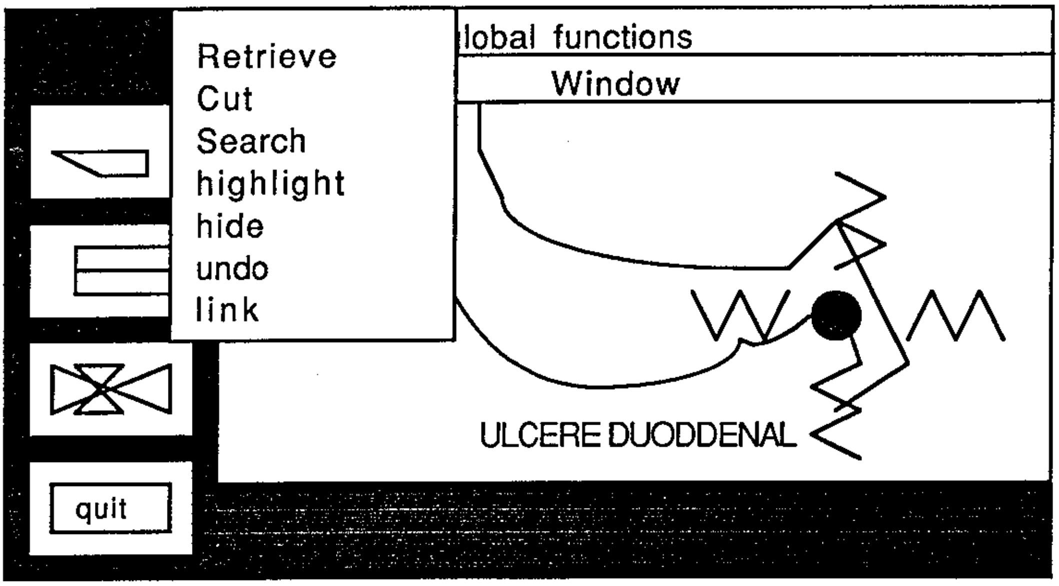
Thus in the STOMACH-DIS and LUNG-DIS relations there is the patient number, the disease number, the disease description, the diagnosis, date and the coordinates of the disease and its width. The picture relation gives the frame number of the patients actual image. Clicking on the SHOW-STRUCT shows the different kinds of objects allowed.

FIGURE A1.12B



If the stomach is selected and then an ulcer is chosen the ulcer icon is then selected and dragged onto the stomach icon.

FIGURE A1.13



EXAMPLES OF USE OF THE ICONIC INTERFACE

The picture can be linked to a particular patient by using the link function. Thus such an iconic interface demonstrates how icons can be practically used within an application.

A1.4 CONCLUSIONS

Direct manipulation is an important attribute of pictorial databases because it enables the user of the system to interact directly with the information contained therein. Thus information can be readily selected and manipulated without an in-depth knowledge of how the data is physically stored within the system. In the case of spatial information, direct manipulation allows users to place the relevant icons in a position that they feel is spatially significant.

Icons are an important attribute of pictorial databases because of the two-dimensionality which enables them to be used in the identification of both spatial and physical information. Within an object oriented interface such icons can be directly manipulated around a screen in order to find a matching spatial arrangement within pictures. However the actual nature of the icon used is dependent on the application area to which it is applied. In the case of restricted domain systems where identical objects are present, then the nature of the icons used has limited importance whereas the spatial configurations of such icons will assume a much greater significance. However in the case of a more generalised system where the nature of the objects within the pictorial database can be more varied then the nature of the icons used assumes much greater importance because here the icon type can represent a class of objects upon which a search of the system can be based. Therefore in this case considerably more care has to be taken concerning the iconic representation used.

APPENDIX 2

ADDITIONS TO THE SPATIAL APPROACH IN SEARCHING PICTORIAL DATABASES

A2.1 INTRODUCTION

Spatial techniques are but one of many attributes which can enable objects within pictures to be located from a pictorial database. Other attributes can also be both visually and textually selected in order to identify a specific picture. These other attributes include:- colour, shape, orientation, size ,number of objects, length, width,perimeter.

Each of these features may be used as an attribute upon which a pictorial database can be searched. However their importance to the user in locating a specific picture depends largely on the the nature of the pictorial database being examined. Therefore in some situations colour may be an excellent attribute upon which to locate a picture from the database, while in other databases shape may be a more important attribute. A short description of each of these attributes now follows.

A2.2 COLOUR

The Young-Helmholtz theory of colour vision suggests there are three different kinds of colour receptors in the eye, each sensitive to a specific portion of the visual spectrum. The first kind of receptor is sensitive to red and its sensitivity declines the more the colour varies from red. The second kind of receptor within the eye is sensitive to green and the third to blue and violet. Again the sensitivity decreases the more that the colour varies from these shades. Therefore to detect an orange yellow colour then the red and green receptors within the eye will be active when observing the colour.

There are three aspects to any colour, its hue, saturation and brightness. Colour vision varies as a function of the wavelength of the stimulus and these various wavelengths are known as hues. Saturation refers to the purity of the wavelengths contributing to the colour sensation, light reflected by different objects is usually multichromatic, therefore the wider the band of wavelength the less saturated is the final colour. Brightness is the amount of black or white mixed in with the hue. The addition of white or black to a hue reduces its purity and brings about a reduction in saturation. Colour is a visual medium and extremely fine differences in shade, brightness and hue can be detected with the human eye.

Colour is an excellent method of visually distinguishing between objects, because of its easy recognition. People implicitly distinguish objects by their colour even when other attributes of objects are as prominent. Colour is therefore often used to describe objects irrespective of the values of the other attributes, for example the red ball, the blue car etc.

Software packages such as Microsoft Windows make considerable use of colour in order to help users rapidly distinguish between different functions of the interface operation.This colourful pictorial element of the interface helps the user to locate the function required without detailed study of the textual instructions involved in the command. Selection of the colour attribute from a menu can be achieved either visually or textually. However where a colour can be selected visually from a palette of available colours on a computer screen, then such as selection will more closely match the user's perceived recall of the colour of an object within a pictorial database. This visual selection of colours will be enhanced if the user has the ability to mix colours from the palette rather than remaining restricted to a simple displayed choice. Textually attempting to describe colours is considerably more difficult. For example there are many shades of red which defy accurate textual description. However visual selection from a colour palette can allow user's greater precision in their query. Unfortunately many objects within pictures contain a number of colours from which it may be difficult to choose a predominant one for the identification of an object from a pictorial database.

A2.3 SIZE

Size is also an important attribute of pictures. The human brain judges certain objects to be of a fixed size and measures surrounding objects by the size of these fixed ones. One such object regarded as being of fixed size is the human body. Therefore in a picture containing a person which has no fixed scale of reference, the human brain associates objects present within the picture in terms of the size of the human frame.

Size can be used quite effectively as a method of pictorial searching. The area of object in question can be accurately determined with the aid of a computer or alternatively by manual techniques. Such an object can often be digitised and a pixel count carried out in order to accurately determine its area value. Having stored this area value it can then be compared with the user's query in order in identify objects from the database having the same value.

One proposed measure of this attribute is relative size. This is the relative area the object of interest occupies within a picture. For example a bird may occupy a large area in one picture while it may be almost indistinguishable in another picture even though in both cases it is the same bird. Under such circumstances the bird can be said to have a larger relative size in the first picture than in the second case. Again as in the case of colour the attribute of relative size can be selected either visually or textually. Visually selecting the attribute value entails either selecting from a choice of predefined values or depicting an area by resizing a given shape. Such a technique should allow users to more accurately specify the size of an object they require rather than being restricted to a given choice. Under these circumstances an icon could be selected from an iconic menu and the representation then scaled to a more appropriate size from which similar objects of the same relative size could be identified from a pictorial database.

Size can also be selected textually by assigning numerical values to increasing values of sizes. For example if the occupation of the entire picture by an object is arbitrarily assigned the size value of ten, then occupation of half the picture by an object would be assigned the value five. Using such a technique relative size could be specified by entering a number representing the relative occupancy of an object within a picture. This value could then be compared with values within a database management system in order to identify objects within pictures occupying the same size on a screen.

A2.4 SHAPE

Objects within a pictorial database do not usually conform to a uniform shape unless artificially constrained to do so. Matching of shapes between the user's specification and the pictorial database is certainly the most difficult of the attributes upon which comparisons can be made. The characteristic of shape can be categorised by such attributes as the number of sides of an object though this approach obviously has its limitations in the case of natural objects such as trees flowers etc.

Matching object shapes falls predominantly in the area of pattern matching, an area still largely in its infancy. When the user specifies a shape on the user-interface then unless an exact match is possible with information within the pictorial database then such comparisons will be difficult. Thus here there is the problem as to what constitutes a matching object. For this reason sketching is rarely allowed within the user-interface for the retrieval of pictorial information because of the difficulty of shape matching between the sketch and the shapes of objects present within the pictorial database.

Because of this difficulty in shape matching, pictorial information systems usually rely upon the use of icons which have a predetermined size and shape and therefore do not directly allow shape comparisons with objects in a database.Thus the problem of comparing user specified shapes with objects in the pictorial database is largely overcome by initially specifying what objects within a pictorial database match to available icons or textual selections from a menu. Using such a technique a database designer initially studies the pictures within a pictorial database classifying objects within these pictures and specifying which objects match with which icons. A user then selects these icons enabling them to locate the corresponding objects within the database. In this way direct comparison between user sketched shapes and objects within pictures is avoided.

A2.5 ORIENTATION OF OBJECTS

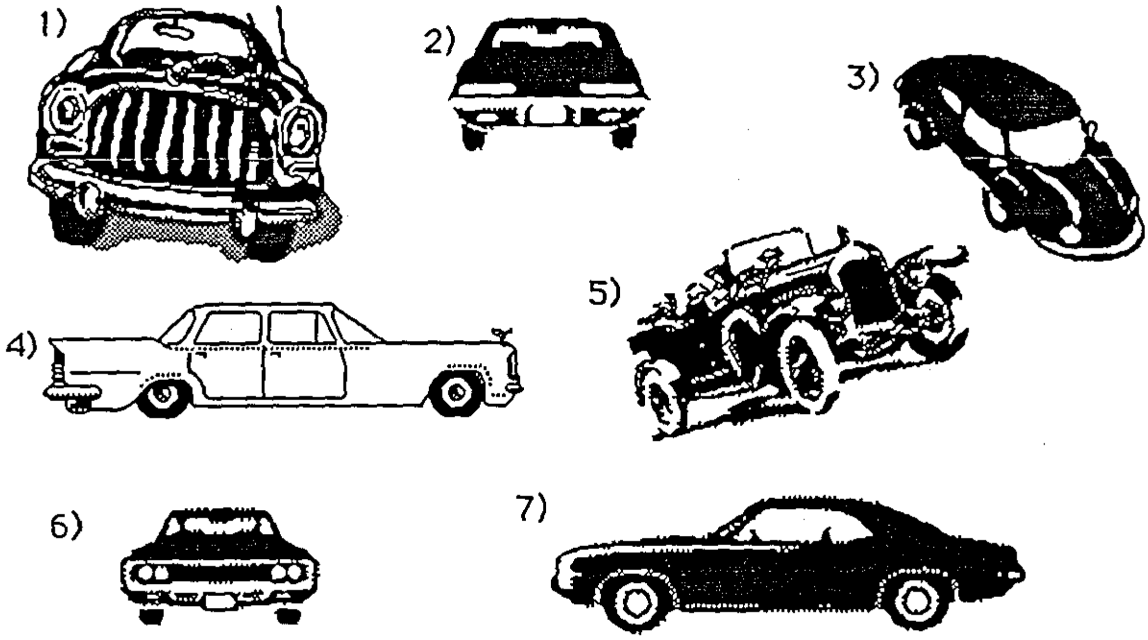
Identical objects can differ in their orientation even when all other attributes such as colour, size and shape remain constant. Many applications such as those involving cellular activity have a strong dependence on the ability of the system to detect changes in object orientation.

The partial rotation of even a simple figure can prevent its recognition provided that the observer is unaware of the rotation. Familiar figures viewed in different orientations often no longer appear to have the same shape. Visual form appears to be based on how parts of a figure are related to one another geometrically. Changing the size and position of objects does not alter the internal geometry of a figure and yet changes in orientation produce a change in appearance. Rock (1974) has proposed that the perceived shape is based on a cognitive process in which the characteristics of the figure are described by the perceptual system. Thus the perception of rotated objects is not only based on the objects internal geometry but takes into account the location of the figures top, bottom and sides.

When multiple shapes are introduced, people find it much harder to detect the correct orientation because the retinal distortion has an adverse effect on recognition. This leads to the proposal that subjects suppress the retinally upright percept and try and substitute a corrected one. However to do this they must visualise how the figure would look if it were rotated until it was upright with respect to themselves. This process of mental rotation requires visualising the entire sequence of angular change and therefore the greater the angular change the greater the difficulty. When assumptions are made about the orientations and shapes of objects then matching between user-specified queries and the information stored on the database becomes considerably easier.

As before, orientation can either be described or depicted and either selected textually or visually. For example a shape can be rotated to more closely correspond to the user's perceived orientation of an object within a picture or selected from available textual values. Thus objects within a pictorial database can have their orientations specified, though in some circumstances what constitutes the 'normal' angle is difficult to determine. The problem here is how to classify the orientation of such objects with respect to the x, y and z axis. For example in the cases below.

FIGURE A2.1



DIFFERENT ORIENTATIONS OF OBJECTS (CARS)

Different orientations of various cars are shown, in order to classify these orientations some convention is needed which both the designer and user of the pictorial database system can readily apply.

One such method might be to classify car 5 as varying in the x-y plane such its orientation could be represented as 30 degrees, 0, 0. . Here the 30 degrees applies to the deviation from the x axis while the zero's apply to the deviation from the the y and z axis. Cars 1,2,4,6 and 7 could be classified as having different z axis angles, for example if cars 1 and car 6 are considered as being the 'normal' angle encountered for a car, then car 4 has a z angle of 90 degrees, car 2 has a z angle of 180 degrees and car 7 has a z angle of 270 degrees. The actual value of the angle therefore depends on what orientation of the object is initially considered as being at angle zero, and where possible there needs to be some convention as to how the objects within the database have their orientation classified. There are many ways in which such orientations could be classified and the variations above demonstrate the complexity of such a problem.

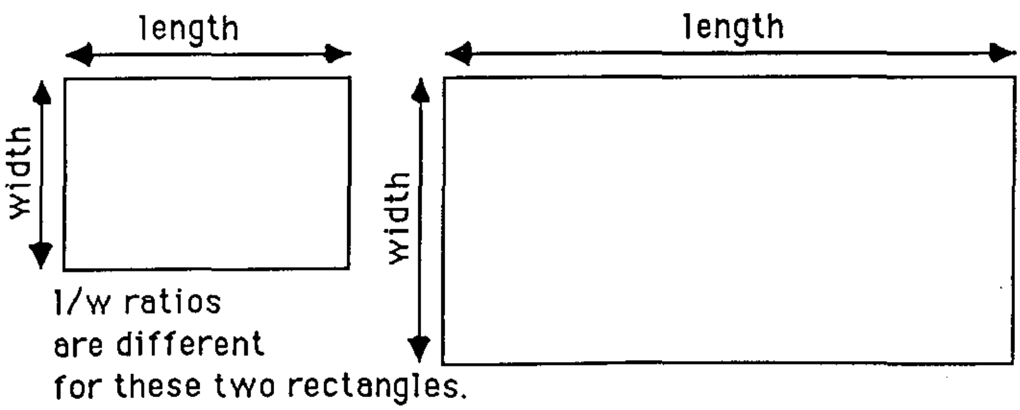
A2.6 NUMBER

Matching the number of objects present within a picture with the user specified query is one of the simplest methods of comparison. Due to its simplicity, this method is often one of the most common attributes used in the classification of objects within pictorial scenes. Database management designers can physically count the number of objects within a picture such as 'five trees' and encode the value of this attribute into a database management system. Users of such systems can then specify the nature of the object required (such as 'tree') either textually or iconically along with specifying the number of such objects (in this case five) that they require within a picture. Matching pictures from a pictorial database containing the same number of similar objects can then be located.

A2.7 LENGTH AND WIDTH

Length and width can be used in order to determine the proportions of the object or objects within a pictorial database. Under these circumstances these attributes are closely related to the minimum enclosed rectangle because of the close correspondence between the size of the rectangle and the values of the corresponding length and width. Users can therefore specify an object's length to width ratio as a means by which to identify an object. Such attributes can be combined with relative size in order to more exactly specify an object from a database. For example:-

FIGURE A2.2



DIFFERING LENGTH AND WIDTH RATIOS

Here depiction of this attribute value by manipulation of a given shape is preferable to description since textually selecting the ration of length to width is in most cases difficult.

A2.8 CONCLUSIONS

There are many object attribute alternatives to visually indicating the spatial location or spatial arrangements of objects within pictures. However which attributes are actually chosen depends in part at least on the nature of the information stored within the pictorial database. A careful consideration of the most appropriate attributes needs to made before choosing the most appropriate ones. For restricted domains containing similar objects one attribute (such as orientation) can often be used to classify nearly all the objects present within a picture. However in the more general case where a variety of objects are present combinations of attributes are usually required to enable the user's of the pictorial database to find the information that they require.

APPENDIX 3

A SIMPLE EXPERIMENT USING SPATIAL SEARCH TECHNIQUES

A3.1 INTRODUCTION

A3.2 DESCRIPTION OF THE SYSTEM

A3.3 THE RESULTS OF THE BALL POSITIONS WITHIN THE BILLIARD DATABASE

A3.4 AN EXAMPLE OF A POSSIBLE USER QUERY

A3.5 A COMPUTERISED SPATIAL SEARCH SYSTEM

A3.1 INTRODUCTION

In order to investigate the possibility of analysing pictorial information in a purely spatial manner, a simple experiment was devised which allowed a pictorial database management system user to select pictures based primarily upon their spatial characteristics. Although the entire system operated manually it could in principle be computerised. The apparatus used to carry out this experiment was a billiards table.

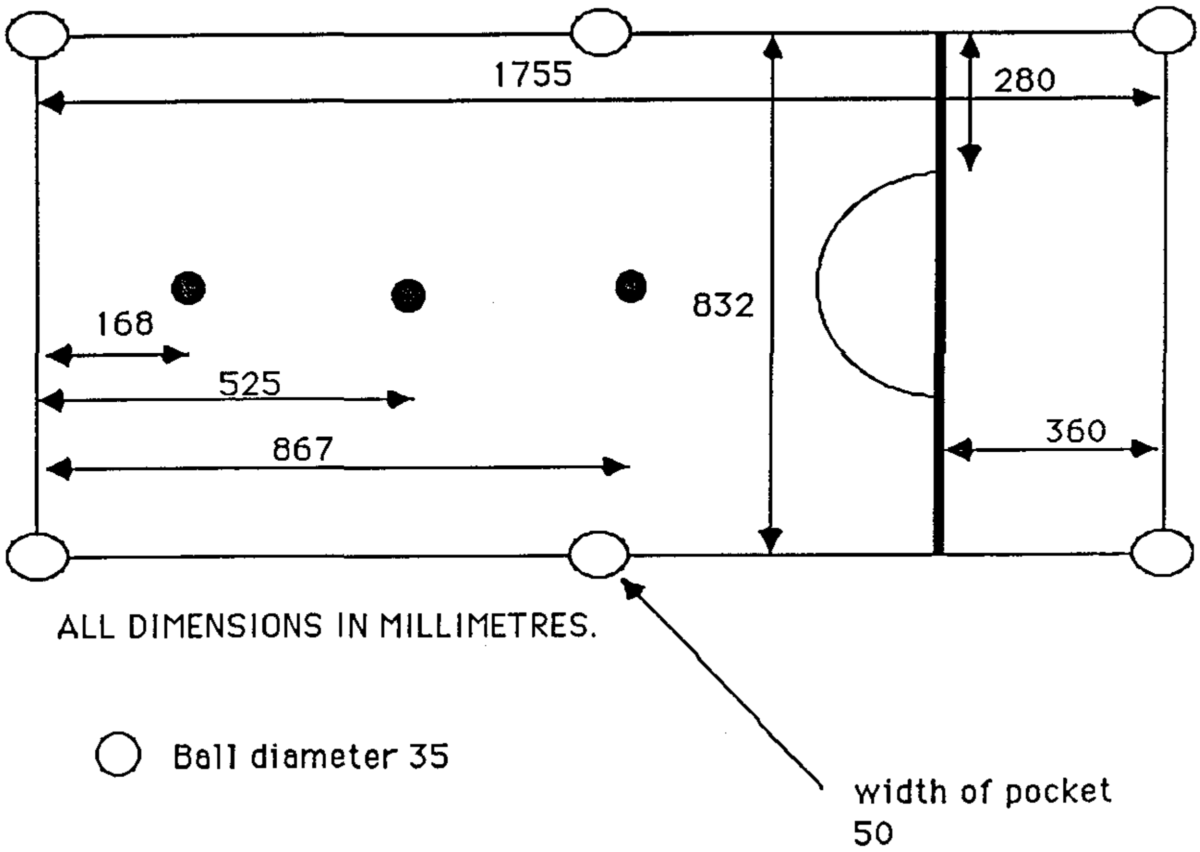
A3.2 DESCRIPTION OF THE SYSTEM

There were several reasons for the selection of an application based upon a billiard table. Firstly the domain was restricted containing only a table and three different coloured balls. Thus while filming the session the area of interest was contained within the four cushions of the billiard table and these boundaries could be clearly seen within the view from the camera. The objects used for this application (billiard balls) were spherical, had uniform colour, no orientation and identical size, therefore encoding of the positions of these balls into a database management system was made simpler. The billiard balls could also be easily differentiated simply by their colour, this ensured that determination of the balls' positions at any one time was easy to record. A sequence of player billiard shots forming the pictorial database was filmed by using a videocamera placed overhead capturing a plan view of the motion of the balls moving around the table.

Linguistic descriptions of the positions of three balls on a table is difficult because both their distance, and spatial relationship to each other need to be taken into account at any one time. In accordance with principles developed in chapter 6 it is suggested that an interactive spatial system could be developed which enabled the user to move icons representing the billiard balls around an interactive user-interface from which matching ball positions could be recalled from the videotaped database of shots. A user of such a system would then be able to identify memorable shots based upon their visual recall of the positions of the balls before or after a shot.

This billiard spatial database system was developed to show that information could be selected from a pictorial database of billiard shots primarily by analysis of the spatial arrangements of the balls, rather than by the analysis of other visual attributes of objects such as size, orientation, colour or shape. A billiard table was specifically chosen in order to eliminate the effect of these other visual attributes as much as possible. The dimensions of the table were taken. These are shown in the diagram below.

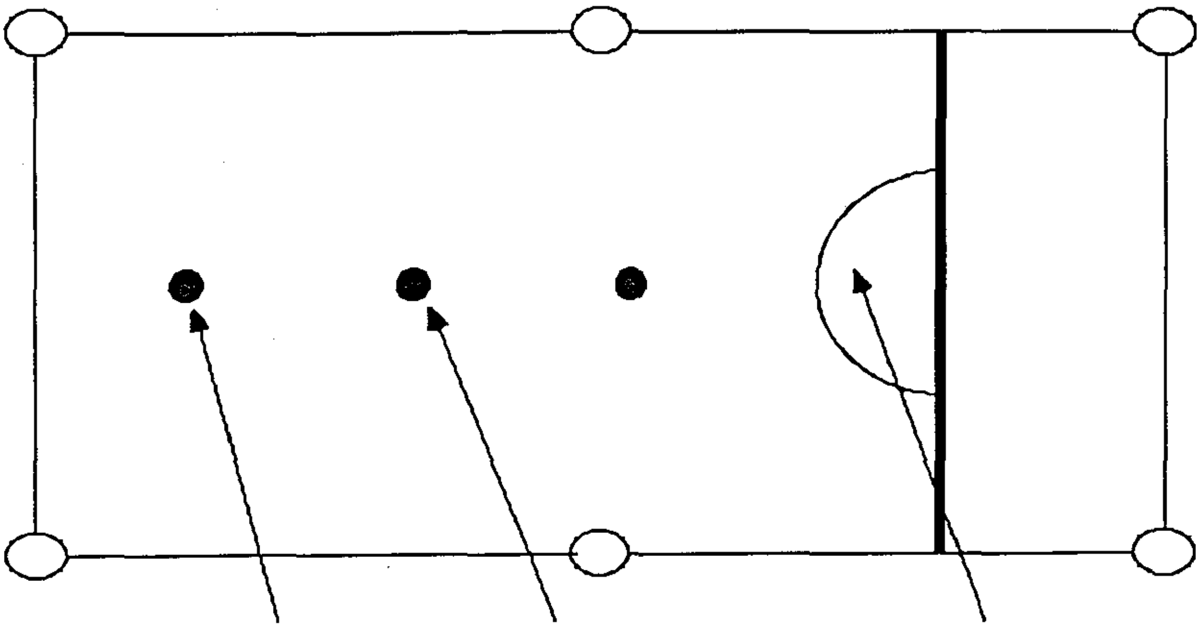
FIGURE A3.1



BILLIARD TABLE DIMENSIONS

Three coloured balls were used, yellow red and black. In order to ensure that three balls were on the table as often as possible, the following variation to the game of billiards was devised. The yellow ball was used as a cue ball and if potted replaced on the table within the semi-circle for the next shot. Both the red ball and the black ball when potted were returned to the first and second spots respectively, as shown below

FIGURE A3.2



POSITIONS OF BILLIARD BALLS ON THE TABLE AFTER HAVING BEEN POTTED

A videocamera was set up overhead giving a plan view of the table and the entire episode of ball movements around the table filmed. Spotlights were used to highlight the table. After each shot the positions of the balls were physically measured from the top left corner of the table. Although time was not an important factor, play was non-stop so as to obtain a continuous videotape of various shots. The table produced by recording the ball positions is shown in section A3.3.

Once all the positions of the ball coordinates for the billiard shots had been recorded the next step was to playback the videotape on a videotape recorder and record the elapsed time at which each shot took place, this was to enable easy reference to shots so that any shot on the videotape could be quickly accessed by knowing the time elapsed since the filming session was started.

For example:-

shot 1 occurred after 4 seconds
shot 2 occurred after 12 seconds
etc for all seventy shots.

In order for users to be able to compare their spatial billiard ball query with the database of possible shots on the videotape, a consultation facility was developed. This enabled the user to move the icons representing the billiard balls around a representation of the table on the computer screens user-interface.

The consultation facility developed was drawn within Apple MacDraw. This was drawn to scale so that users interactively manipulating the representations of the balls around the user-interface with a mouse did not encounter any error due to scale differences. The object of this exercise was to try to find a picture containing the same spatial arrangement of billiard balls from the videotape database of shots to the user's spatial query. If it could be show that, in principle, placing the icons of the balls on the user-interface found a closely matching spatial arrangement of balls to that within a frame of the videotape, then it would seem feasible that depicting the spatial arrangement of the objects would appear to be a reasonable method of searching a pictorial database in cases where pictures were difficult to locate using conventional textual query languages.

A3.3 THE RESULTS OF THE BALL POSITIONS WITHIN THE BILLIARD DATABASE

FIGURE A3.3

SHOT	YELLOW BALL X	Y	RED BALL X	Y	BLACK BALL X	Y
1	32	170	45	162	59	111
2	59	111	71	170	80	77
3	15	100	63	109	80	77
4	48	163	CLP		80	77
5	78	99	7	19	80	77
6	76	81	7	19	50	125
7	20	24	34	27	50	125
8	75	164	55	11	50	125
9	53	125	21	42	47	138
10	70	157	21	42	77	84
11	61	65	21	42	73	51
12	14	79	21	42	30	15
13	15	7	36	40	45	15
14	76	122	36	40	84	11
15	65	49	36	40	39	24
16	78	135	36	40	61	21
17	19	113	36	40	12	49

BALL COORDINATES (SHOT 1-17)

FIGURE A.3.4

SHOT	YELLOW BALL X	YELLOW BALL Y	RED BALL X	RED BALL Y	BLACK BALL X	BLACK BALL Y
18	37	33	38	29	12	49
19	48	40	49	18	12	49
20	16	43	49	18	84	66
21	43	42	RBP		84	66
22	81	66	RBP		17	116
23	76	152	22	63	17	116
24	11	105	22	63	80	59
25	21	28	22	63	81	72
26	46	28	72	63	26	156
27	76	120	12	151	26	156
28	84	123	17	141	26	156
29	22	127	49	165	48	151
30	37	158	49	165	78	123
31	51	153	68	149	78	125
32	72	34	10	150	83	118
33	36	158	10	150	28	113
34	54	97	10	150	22	96

BALL COORDINATES (SHOT 18-34)

FIGURE A3.5

SHOT	YELLOW BALL X	YELLOW BALL Y	RED BALL X	RED BALL Y	BLACK BALL X	BLACK BALL Y
35	36	169	3	178	LMP	
36	16	153	TLP		LMP	
37	53	80	TLP		54	37
38	31	138	62	44	54	37
39	18	41	62	44	71	152
40	58	153	44	33	71	152
41	69	126	44	33	40	142
42	56	85	44	33	72	83
43	21	52	12	40	74	119
44	48	83	38	20	74	119
45	79	149	38	20	32	148
46	59	102	38	20	48	167
47	11	9	38	20	34	152
48	60	106	15	37	34	152
49	19	24	26	104	34	152
50	82	177	28	63	34	152
51	15	177	28	63	23	89
52	29	64	60	30	14	77

BALL COORDINATES (SHOT 35-52)

FIGURE A3.6

SHOT	YELLOW BALL X	YELLOW BALL Y	RED BALL X	RED BALL Y	BLACK BALL X	BLACK BALL Y
53	16	92	34	143	16	77
54	42	51	32	144	28	142
55	76	177	6	144	34	120
56	6	141	53	132	34	120
57	34	68	53	132	51	106
58	29	144	53	132	79	144
59	79	143	19	93	79	144
60	18	52	19	93	78	154
61	36	157	82	155	83	177
62	51	161	61	136	83	177
63	80	123	76	91	83	172
64	87	178	76	91	78	160
65	79	165	76	91	66	79
66	19	13	76	91	59	79
67	71	155	76	91	77	168
68	67	28	76	91	77	168
69	44	18	11	120	77	168
70	51	128	60	118	77	168

BALL COORDINATES (SHOT 53-70)

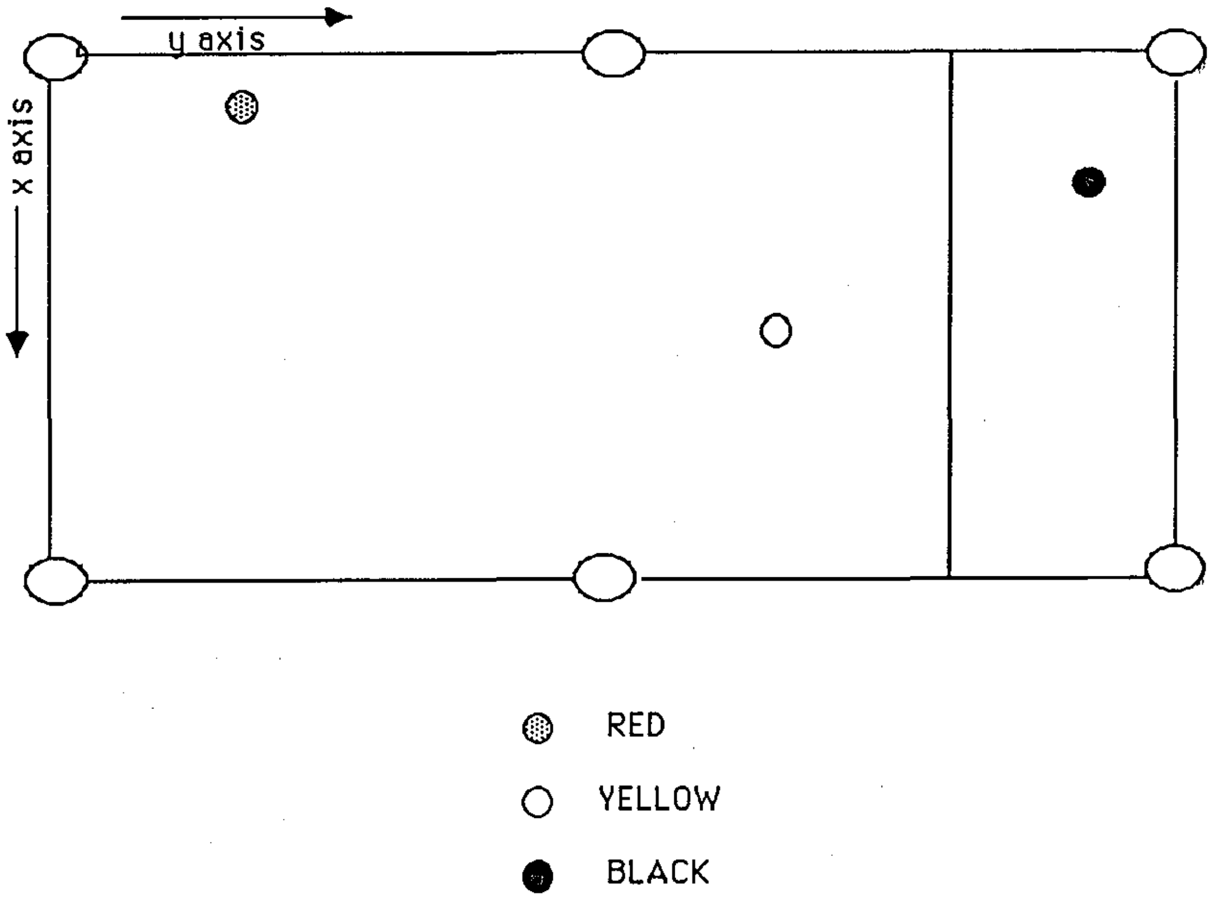
CRP CENTRE RIGHT POCKET   CLP CENTRE LEFT POCKET

TLP TOP LEFT POCKET   RBP RIGHT BOTTOM POCKET

A3.4 <u>AN EXAMPLE OF A POSSIBLE USER QUERY</u>

In the diagram below the user-interface is shown with a depicted query. Here the user is looking for the spatial layout of the billiard balls shown in the diagram.

FIGURE A3.7



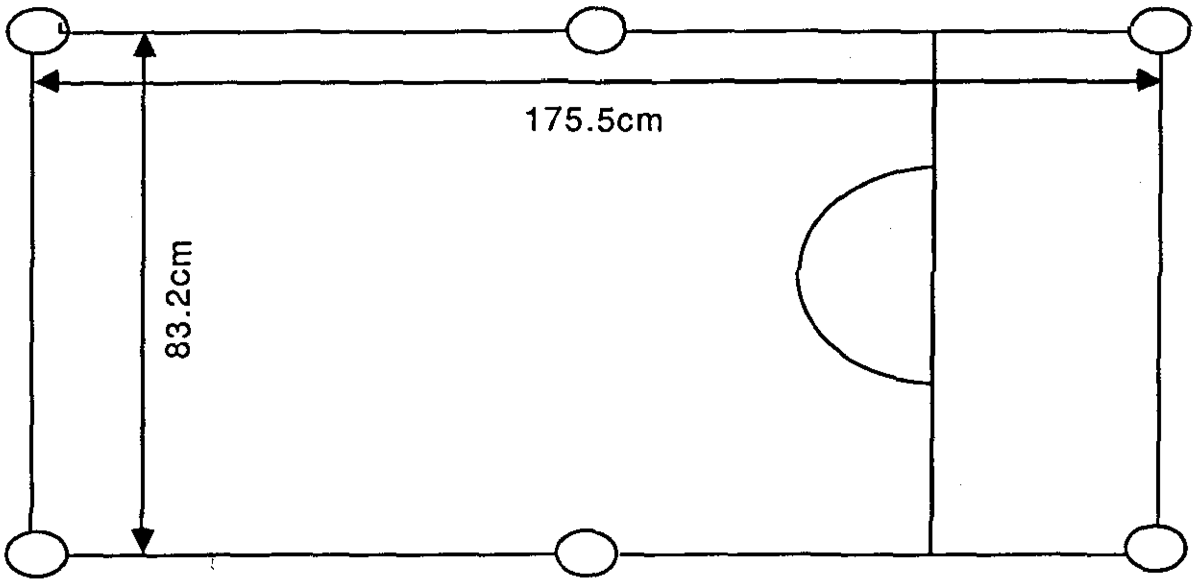
USER'S SPATIAL BILLIARD QUERY

The user has moved the icons of the balls to the positions above using a mouse in an attempt to find a matching picture from the videotape database of billiard shots. In order to compare the user query with each picture from the videotape, several calculations are required. First the coordinates of the iconic representations of the balls need to be calculated from the user's query above. This is achieved by manually taking a printout of the picture above and physically measuring in centimetres the positions of the balls. A more sophisticated computerised system should be able to directly determine such coordinates from the user-interface. Each set of coordinates of shots within the videotaped database is then compared with the coordinates of the balls from the users spatial query above. Since the balls' coordinates in figure A3.7 were

Yellow ball	x 3.7cm	Red ball	x 0.7cm	Black ball	x 1.8cm
	y 9.7cm		y 2.6cm		y 13.9cm

The actual size of the table had been measured and found to have the following dimensions.

FIGURE A3.8



BILLIARD TABLE DIMENSIONS

Then from these table coordinates the corresponding positions of the user's query to the balls on the actual table can be calculated by simple ratios. These calculations are shown below.

YELLOW BALL COORDINATES

x is 3.7 cm

since table width is 83.2 cm and this is equivalent to 7 cm on the user's diagram then this represents

$$\frac{3.7 * 83.2}{7} = 43.9 \text{ cm on the billiard table}$$

y is 9.7 cm

since table length is 175.5 cm and this is equivalent to 15 cm on the user's diagram, then this represents

$$\frac{9.7 * 175.5}{15} = 113.39 \text{ cm on the billiard table}$$

RED BALL COORDINATES

x is 0.7 cm

since table width is 83.2 cm and this is equivalent to 7 cm on the user's diagram then this represents

$$\frac{0.7 * 83.2}{7} = 8.32 \text{ cm on the billiard table}$$

y is 2.6 cm

since table length is 175.5 cm and this is equivalent to 15 cm on the user's diagram then this represents

$$\frac{2.6 * 175.5}{15} = 30.42 \text{ cm on the billiard table}$$

BLACK BALL COORDINATES

x is 1.8 cm

since table width is 83.2 cm and this is equivalent to 7 cm on the user's diagram then this represents

$$\frac{1.8 * 83.2}{7} = 21.3 \text{ cm on the billiard table}$$

y is 13.9 cm

since table length is 175.5 cm and this is equivalent to 15 cm on the user's diagram then this represents

$$\frac{13.9 * 175.5}{15} = 162.63 \text{ cm on the billiard table}$$

Now since the actual positions of the balls for each shot was known (having been measured during the videotape filming) then the distances between the balls in the user's query and the balls during each of the shots during filming can be calculated. For example in shot one the positions of the balls was recorded as being

Yellow ball x 32cm	Red ball x 45cm	Black ball x 59cm
y 170cm	y 162cm	y 111cm

Therefore a comparison can be made between the positions of the balls in the user's query with its own colour ball for each shot of the videotape film.

Thus by pythageros theorem:-

For shot 1

the distance between the yellow ball for shot 1 and the user's yellow ball query is given by

$$\sqrt{(43.9-32)^2 + (113.49-170)^2}$$

$$= 14.07\text{cm}$$

the distance between the red ball for shot 1 and the user's red ball query is given by

$$\sqrt{(8.32-45)^2 + (30.42-162)^2}$$

$$= 136.57\text{cm}$$

the distance between the black ball for shot 1 and the user's black ball query is given by

$$\sqrt{(21.3-59)^2 + (162.63-111)^2}$$

$$=63.92\text{cm}$$

Therefore these three coordinates are a measure of how well shot number one compares with the users spatial query specified by placing icons on the user-interface representing the billiard table. In order to explain how the user's query was compared with the information stored within the database a discussion of two possible forms of analysis ensues.

If subject A has quite a good memory for the general spatial arrangements of the objects presented but does not remember the exact position of each then as a result, subject A places all the objects requested in close proximity to their original location in the interface but none in the exact position. Subject B on the other hand does not remember the spatial location of any of the objects presented except one. So subject B randomly places all the objects in relation to the spatial arrangement in which they had appeared, except the one which is placed in its exact position. Use of a measure of percentage of objects memorised in their exact location as is customarily done (Naveh-Benjamin 1988) will result in attributing better memory for spatial location to subject B which can be quite misleading in determining which subject best remembers the locations of specified objects. A performance measure should be available for expressing the advantage of subject A in placing most objects closer to their original spatial location. One such possible measure by which the absolute positions of objects can be compared is the mean absolute deviation of subject's memory for the spatial location of the objects as specified in the user-interface from their original arrangement within a picture. This measure provides information about how close a subject's memory of the positions of the objects is to the original spatial configuration of objects within the pictures. Thus in order to find a more representative value of comparison this mean absolute deviation was calculated for all of the seventy shots.

Thus for shot 1 the mean deviation is

$$\frac{14.07 + 136.57 + 63.92}{3} = 71.52$$

Therefore if each shot is analysed in the same way, positions added and divided by the number of balls (in this case three) then the minimum value of these numbers representing the mean absolute deviation for each shot will correspond to the closest matching picture within the pictorial database to the user's query specified earlier. These calculations were carried out for all seventy shots and the resulting mean deviations are shown in figure A3.9 to figure A3.13.

FIGURE A3.9

SHOT	1	71.5
SHOT	2	90.3
SHOT	3	77.3
SHOT	4	70.3
SHOT	5	50.8
SHOT	6	34.9
SHOT	7	55.3
SHOT	8	52.4
SHOT	9	22.5
SHOT	10	51.4
SHOT	11	63.9
SHOT	12	70.2
SHOT	13	96.4
SHOT	14	75.5
SHOT	15	79.0
SHOT	16	72.2
SHOT	17	104.8
SHOT	18	87.4
SHOT	19	89.3
SHOT	20	74.6

FIGURE A3.10

SHOT 21	72.5
SHOT 22	62.6
SHOT 23	44.1
SHOT 24	62.8
SHOT 25	77.4
SHOT 26	55.7
SHOT 27	64.1
SHOT 28	53.4
SHOT 29	95.1
SHOT 30	84.9
SHOT 31	80.7
SHOT 32	83.5
SHOT 33	71.6
SHOT 34	68.5
SHOT 35	103.9
SHOT 36	75.5
SHOT 37	73.8
SHOT 38	70.8
SHOT 39	87.4
SHOT 40	80.1

FIGURE A3.11

SHOT	41	30.5
SHOT	42	53.7
SHOT	43	48.1
SHOT	44	45.5
SHOT	45	33.1
SHOT	46	25.8
SHOT	47	52.5
SHOT	48	14.6
SHOT	49	74.0
SHOT	50	42.9
SHOT	51	59.8
SHOT	52	63.1
SHOT	53	79.2
SHOT	54	66.7
SHOT	55	76.6
SHOT	56	67.4
SHOT	57	73.8
SHOT	58	68.5
SHOT	59	67.6
SHOT	60	62.5

FIGURE A3.12

SHOT	61	84.1
SHOT	62	76.5
SHOT	63	63.5
SHOT	64	105.9
SHOT	65	82.7
SHOT	66	95.3
SHOT	67	65.4
SHOT	68	78.4
SHOT	69	101.4
SHOT	70	79.0

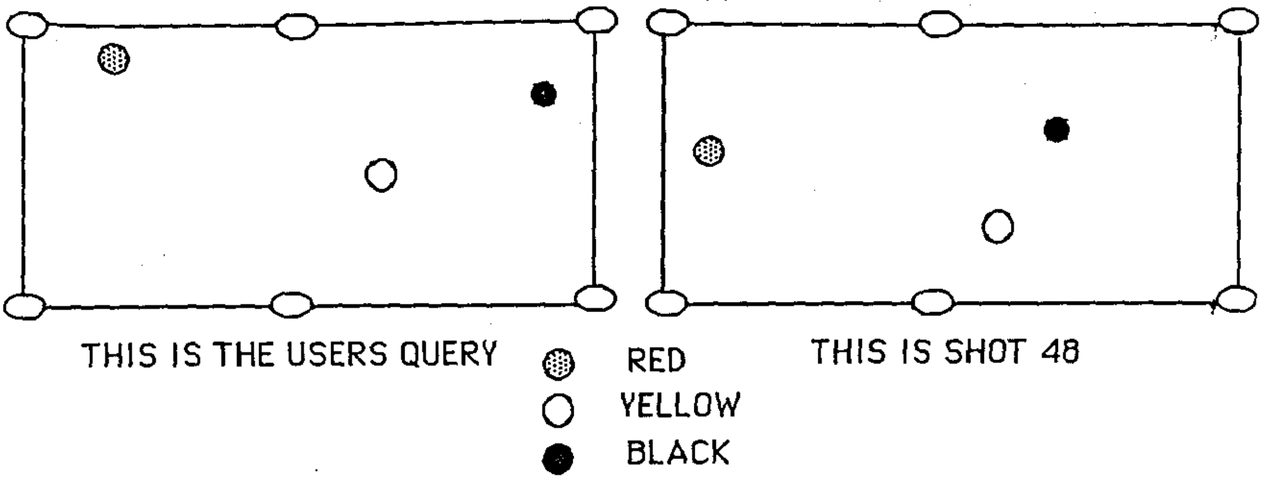
FIGURE A3.13

SHOT	48	14.6
SHOT	9	22.5
SHOT	46	25.8
SHOT	41	30.5
SHOT	45	33.1
SHOT	6	34.9

FIRST SIX SHOTS WITH THE SMALLEST MEAN DEVIATIONS FROM THE USER QUERY

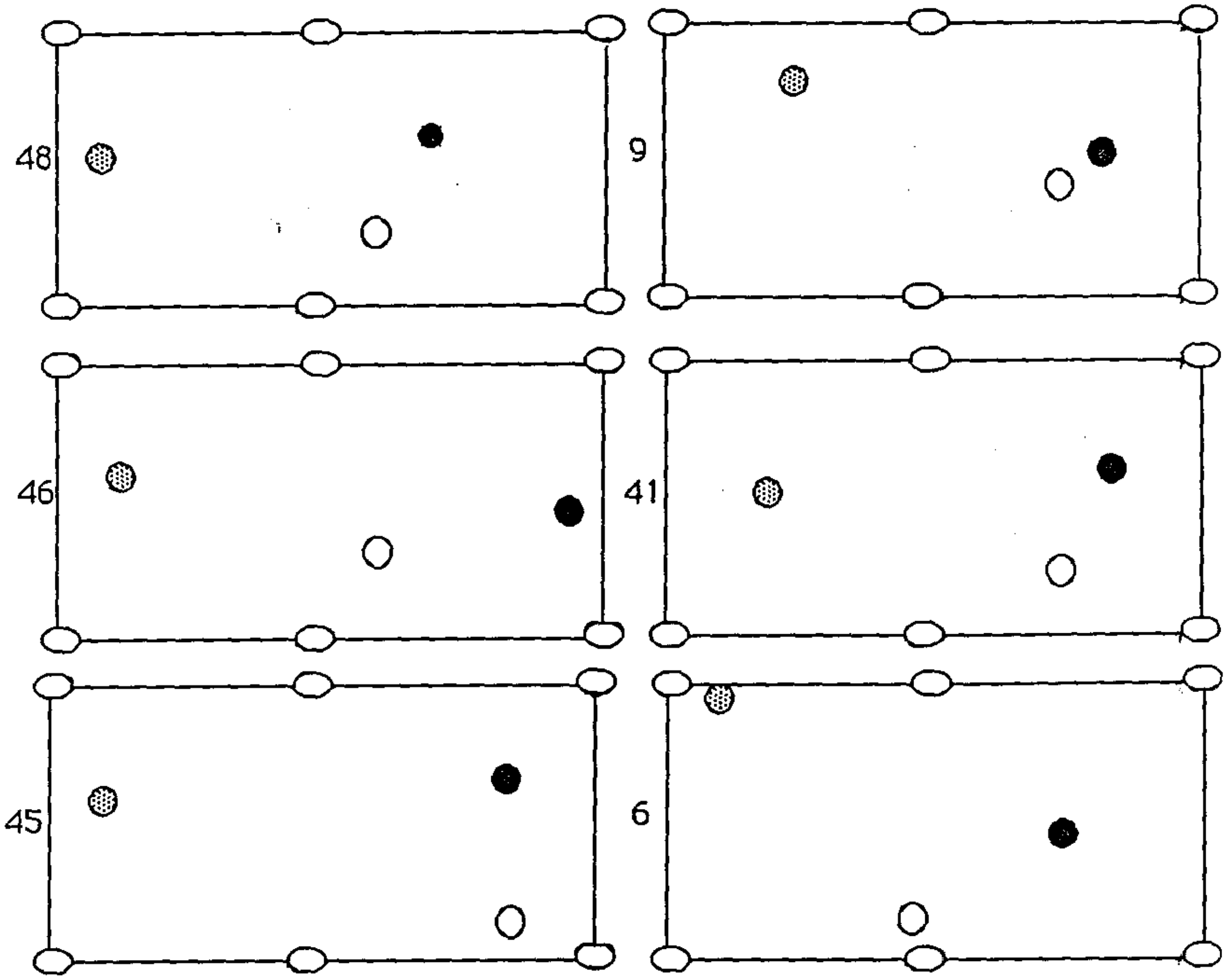
The smallest value of the mean deviation from the user's query was therefore found to be that of shot 48. This shot is represented pictorially in figure A3.9 next to the user's original depicted spatial query.

FIGURE A3.14



COMPARISON OF USERS QUERY WITH SHOT 48

FIGURE A3.15



FIRST SIX SHOTS WITH THE SMALLEST MEAN DEVIATIONS

In comparing shot 48 with the user's spatial query it can be seen that there is certainly some similarity between the two pictures, both having the same overall positional relationship between each of the respective balls. However since the pictorial database present on the videotape consisted of a limited number of shots (70 in all), it is to be expected that there is only a reasonable degree of match in the absolute positions of the balls. If a larger number of shots had been filmed it would be expected that the likelihood of a closer match would be more probable. However the principal of spatial search still applies in that the above user query and result closest matching picture suggests by their closeness of visual match, that information within a pictorial database can be queried by depicting visual queries rather than simply by specification of such queries textually.

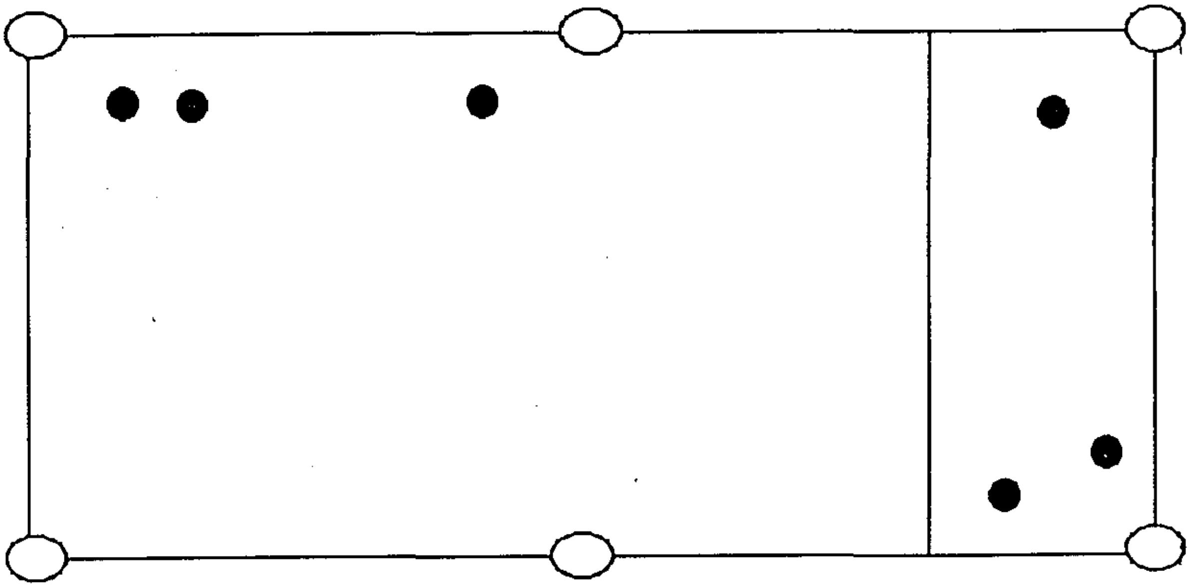
A3.5 A COMPUTERISED SPATIAL SEARCH SYSTEM

In order to further demonstrate the principle of absolute spatial search a computerised application was developed using Apple HyperCard. This consisted of twenty different drawn pictures constituting the pictorial database, each picture containing six identical of balls arranged randomly on a computerised graphical representation of a billiard table. One such picture is shown in figure A3.16. An interface was constructed identical in appearance to the picture below, allowing the user to move the positions of the balls around the screen by mouse control. The system has been constructed so that users do not require a detailed knowledge of the structure of the pictorial database or the technique by which the ball's position within each picture is stored. The user would then have the ability to move between different balls simply by selecting them with the mouse.

In common with many spatial database management systems, the system developed would allow users to specify positional queries concurrently. This is in contrast to textual query languages in which successive queries must be posed to the database. Users would be asked to depict the spatial arrangement of ball in a top to bottom fashion in accordance with Wilton and File's finding that for identical objects, subjects appear to remember objects in that form.

The resulting configuration of user arranged balls on the interface would then be compared with the ball positions in each of the twenty pictures in order to find the closest matching picture. This is achieved by storing the positions of the balls within the twenty pictures in an Oracle relational database and comparing each ball position on the interface with each of the positions of the balls in the relational database. The Oracle database system was designed such that new pictures could be easily constructed and the balls' positions within these pictures entered into the fields of the relational database. Since all the balls within the current twenty pictures are physically identical, the only possible method by which the user would be able to identify a chosen picture is by the specification of the spatial positions of the balls on the interface, (other attributes such as size, shape and colour of the balls being the same).

FIGURE A3.16



DRAWN PICTURE OF IDENTICAL BALLS SPATIALLY ARRANGED ON A BILLIARD TABLE

Psychological evidence has suggested that users are able to remember spatial arrangements of identical objects and that these objects are stored in memory as higher-order units.Thus it is suggested that this computerised spatial system should allow the user to identify a picture containing different spatial arrangements of identical objects from a pictorial database.

It is suggested that his HyperCard system should demonstrate that the spatial method of search can be used as a technique by which specific pictures containing identical objects can be found. However the domain here is very restricted in that all objects are identical consisting of only simple iconic representations of balls.

Although the absolute deviation measure takes into consideration absolute characteristics of memory performance, it does not provide a direct account of relational information. To demonstrate this point, assume that subject C knows the general location of each of the objects in respect to their original positions but does not completely remember the spatial relations of the objects to each other (eg whether one is to the left of or above the other). In contrast subject D does not completely remember the general location of each of the objects but has a good idea about the spatial relations between such objects. For example subject D remembers that some objects appeared to the left of others and others appeared above. Using either of the two aforementioned measures in providing information about accuracy of performance in an absolute sense (percentage of balls positions remembered or approximate absolute location within a picture) could lead to the conclusion that subject C has better memory for spatial location information. This is a misleading conclusion because in terms of information about relative spatial position subject D definitely has an advantage. Hence the need arises for another measure to provide information about memory for spatial relations between the objects. Two such measures are suggested in describing each subjects performance. One is memory for relative position along the horizontal dimension and the other is memory for relative position along the vertical dimension. Similar techniques to both of these measures have been described earlier in chapter 6 although in this case identical objects rather than objects differing visually are used. It is suggested that all these measures can be taken into consideration when examining spatial data.

APPENDIX 4

THE HYPERCARD DATABASE MANAGEMENT SYSTEM

A4.1 USING THE HYPERCARD DATABASE MANAGEMENT SYSTEM

A4.1.1 Demonstrations of the System

A4.1.2 The Van Gogh Iconic Menu

A4.1.3 The Van Gogh Textual Menu

A4.1.4 The Relational Database

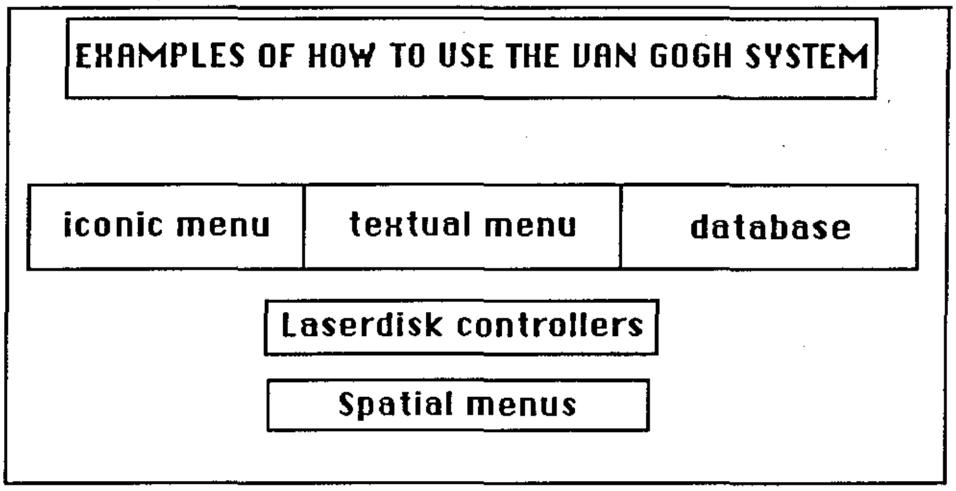
A4.1.5 The Visual Depiction of Queries

A4.1.6 The Laserdisk and its Controller

A4.1 USING THE HYPERCARD DATABASE MANAGEMENT SYSTEM

After initially logging in to the HyperCard database management system with the username and password, users then select from one of six choices on an initial menu, these are:- examples of how to use the Van Gogh system, iconic selection, textual selection, relational database, laserdisk controller and spatial menus. Having selected one of these choices users then attempt to locate the picture (or pictures) that they are interested in, either by selection of icons representing objects from a menu, textual selection of object names from a menu, selection of object attributes by textual entry into a relational database, visually depicting the values of object attributes such as size and orientation, or by browsing through the laserdisk by using one of the two on-screen laserdisk controllers provided. At any point within the database management system users have the ability either to return to the initial menu shown in figure A4.1 or in many cases go directly to other parts of the database management system, providing the ability to both easily and quickly locate information that they require. The initial menu upon entering the Van Gogh database management system is shown in figure A4.1.

FIGURE A4.1



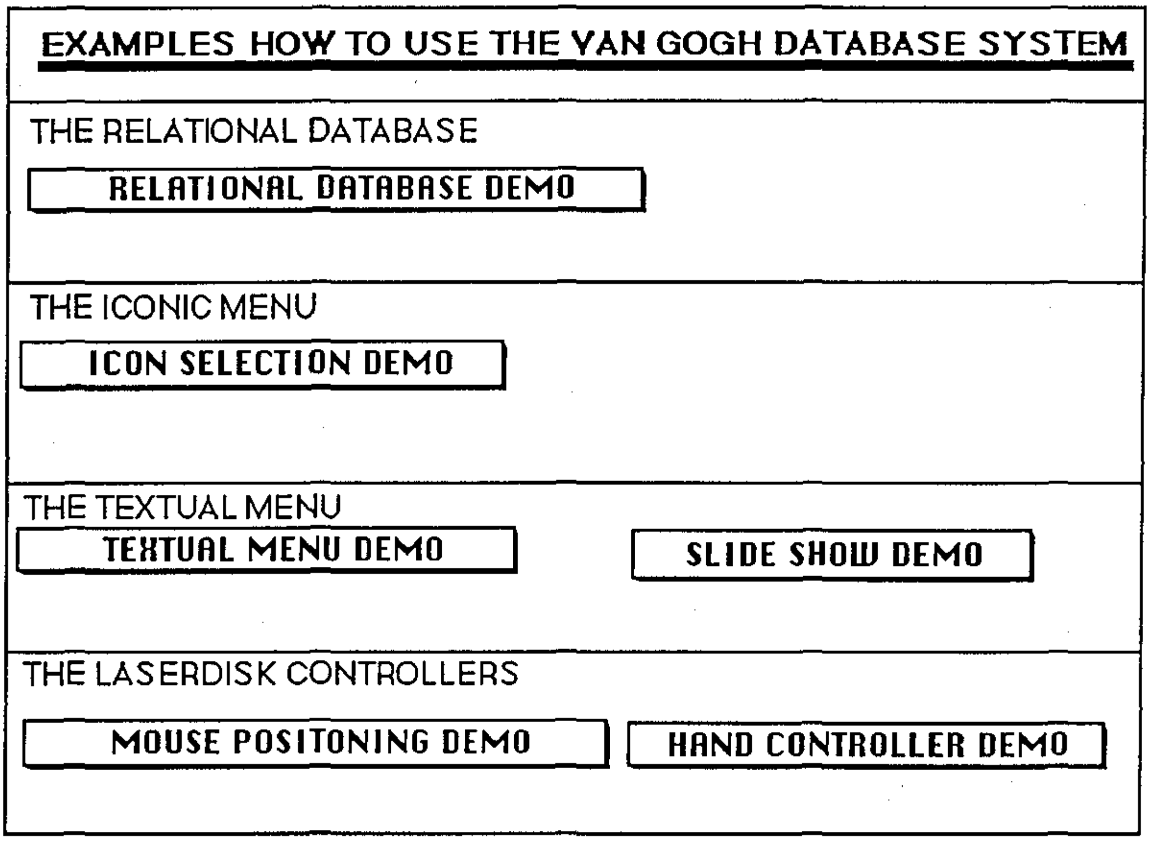
INITIAL MENU FOR THE VAN GOGH DATABASE SYSTEM

Sound in the form of speech has also been incorporated into the Van Gogh database system by using a device called a Macrecorder, which attached comments and commentary to objects and attributes of the pictorial database, however since such a facility occupied a large disk volume, this facility was added for demonstration purposes only. Given sufficient disk volume, sound could easily be attached to any aspect of the database either as an alternative to, or in addition to, displaying text or graphics on screen.

A4.1.1 Demonstrations of the System

On clicking the **EXAMPLES** button from the initial menu above the following screen is displayed

FIGURE A4.2



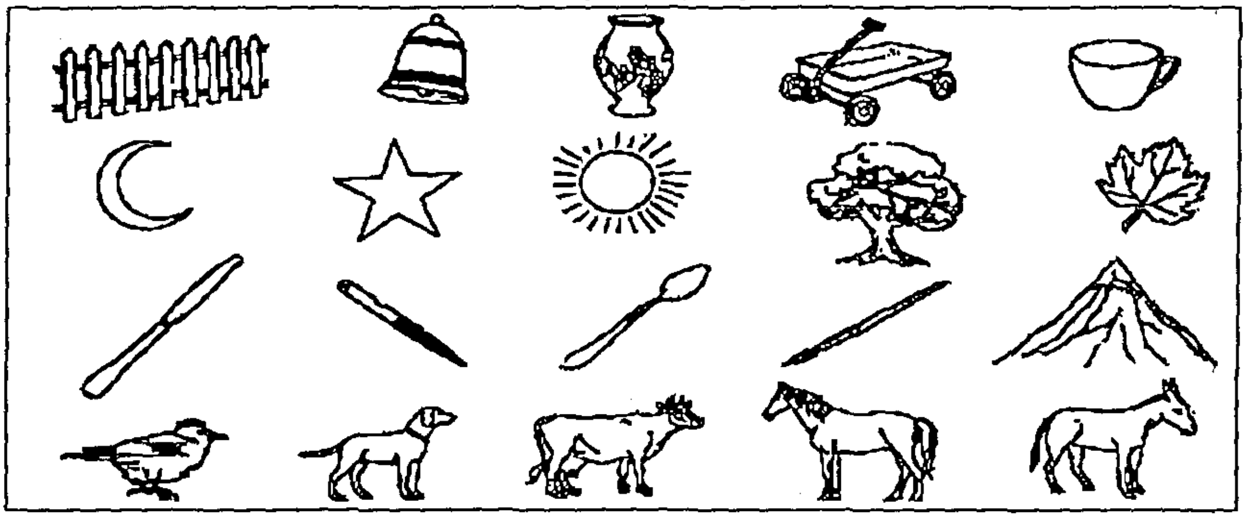
VAN GOGH DEMONSTRATION INTERFACE

Clicking on any of the **DEMO** buttons above automatically displays a recorded series of computer screens which are shown in succession to the user, demonstrating how each of the various aspects of the Van Gogh database management system is used.

A4.1.2 The Van Gogh Iconic Menu

One screen of the Van Gogh database management system's iconic menu is shown in figure A4.3. Each of the icons shown represents objects contained within pictures in the laserdisk's pictorial database of Van Gogh paintings. In cases where there is only one instance of the object present within the database then selection of the appropriate icon will directly display the picture from the laserdisk containing that object. In cases where there are many objects within pictures which can match to a chosen icon, then in order to reduce the number of possible matches more detailed descriptions can be selected from the iconic or textual menus, or alternatively physical attribute values of the object required can be depicted or chosen from the Oracle relational database.

FIGURE A4.3

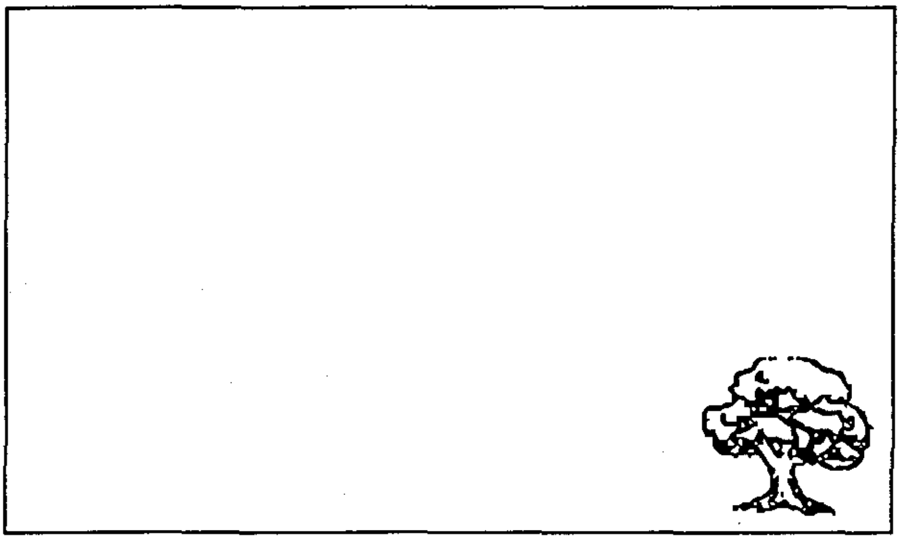


ICONIC MENU FOR THE VAN GOGH DATABASE SYSTEM

For example since the Van Gogh laserdisk contains many pictures of trees, then the selection of the tree icon from the iconic menu above allows the user to further textually specify a more detailed description of the type of tree in question, such as Oak, Beech, Rosewood, etc from the textual menu. Alternatively if the name of the tree required is not known then its physical attributes such as relative size, colour or absolute spatial location within the picture can be chosen from the relational database or alternatively values of these physical attributes can be visually depicted. A visual depiction might for example specify that the tree is inclined at an angle of 30 degrees to the horizontal. Thus the iconic menu allows users to first select specific pictures based upon the physical appearance of objects contained within them, and then further select more detailed information either by choosing items from a textual menu, relational database or by depicting values on an interactive user-interface.

The pictorial database management system has been designed so that the selection of an icon by the user enables the database management system to enter a textual description of the object the icon represents into a SQL select query statement of the Oracle relational database. For example choosing the icon of a tree from the iconic menu above directly enters the word 'TREE' into the object name field of the relational database. This enables the user to select all encoded pictures containing trees from the laserdisk, which are then displayed. A facility has also been developed in which the user can place an icon on an interactive screen enabling matching of both name and spatial location of that icon to information stored within the laserdisk's pictorial database. Therefore selecting an icon of a tree (say) and placing it in the bottom right hand corner of an interactive screen will find all pictures within the laserdisk containing trees which are located in the bottom right hand corner. This icon placement on the user-interface is shown in figure A4.4.

FIGURE A4.4



ICON OF A TREE PLACED WITHIN AN INTERACTIVE INTERFACE

A4.1.3 <u>The Van Gogh Textual Menu</u>

For completeness textual selection from a menu is also available within the Van Gogh database management system. Here the objects present within the pictorial database are textually described, as are their logical relationships to their composite objects. For example a house (which can be selected, finding all houses in the pictorial database) is logically related to doors, windows etc, these doors and windows can also be used as part of the search procedure by which to identify specific pictures. Thus the user can navigate through the available textual menus identifying information that they require. Also provided within the textual menus are the headings 'TYPES' which allows the user to specify the type of object that is required. Each item is selected from the textual menu by simply clicking the mouse on the appropriate descriptive word. The concepts behind the design of the textual menu hierarchy have been taken from Economopoulos and Lochovsky's paper entitled: A system for Managing Image Data (1983). Using these principles a typical textual screen menu within the Van Gogh system is shown in figure A4.5.

FIGURE A4.5

SPATIAL MENUS

ICONIC MENU	TEXTUAL MENU	DATABASE	DISK CONTROLLER

OBJECT	HAS-COMPONENTS	IS-PART-OF	TYPES
PEOPLE	HANDS TORSO	CROWD	MAN WOMAN BABY

Previous Menu

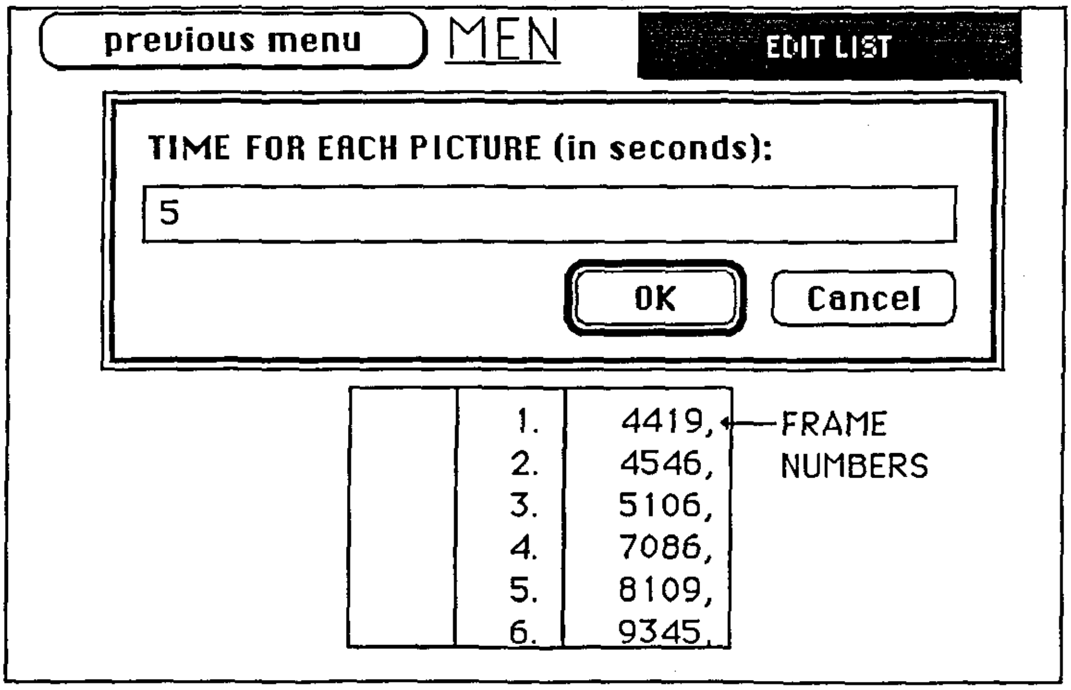
EXAMPLE OF A TEXTUAL SCREEN MENU FOR THE VAN GOGH DATABASE

Therefore if the user of the system initially selects **PEOPLE** from a previous menu then the above table appears, enabling the user to be more specific in his/her query as to the nature of the people or the components of the people that they are interested in. Objects within this menu may in turn themselves be logically linked to yet further menus containing yet more detailed objects. Where the selection of an object from the textual menu has only one matching picture then the database management system directly displays that picture without recourse to the relational database. If the selection of an item from the textual menu does not lead to a single picture then the object name selected is directly entered into the relational database from which all pictures containing objects which match the user criterion can be displayed. Where applicable objects were described in terms of generalised cylinders in accordance with Marr's theories of object decomposition (see chapter 5).

Having identified the name of the object required from the textual menus in cases where there are more than one matching object within the pictorial database, then other attributes such as the absolute spatial location of the object can be specified either by direct numerical entry into the relational database or by indicating where the object is spatially located within a picture on the appropriate interface. This places the numeric values of the attributes chosen directly into the relational database from which matching objects within the pictorial database having similar visual attributes can be located.

In addition to allowing a specific picture to be selected from the textual menu, an additional facility has been developed which allows the user to quickly see all the pictures containing the object required, for example all pictures containing trees. Here the user is asked for the time each picture should be displayed on the television screen, and having entered the answer in seconds the database management system then displays to the user each picture containing the object. During this slide show the user can stop the sequence at any time in order for a more detailed study of a specific picture. Pictures within the sequence can also be easily added or removed from the slide show displaying a different sequence of pictures to the user in situations where the picture content of the pictorial database needs to be changed. The slide show interface is shown below.

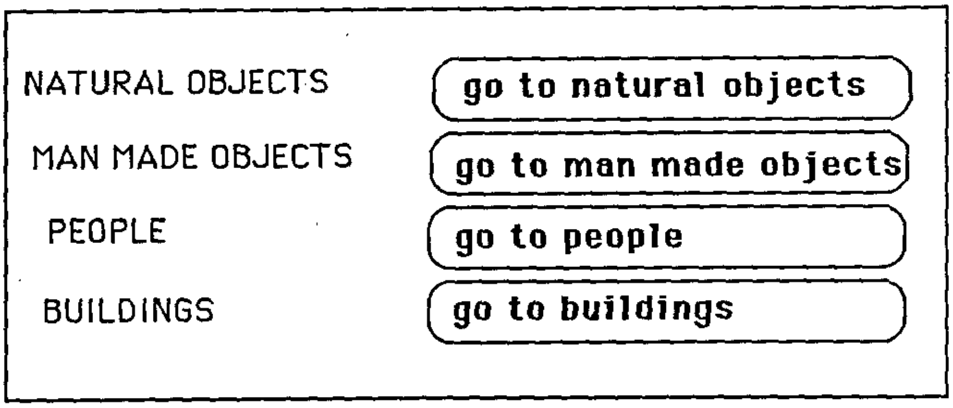
FIGURE A4.6



SLIDE SHOW INTERFACE

Practically the Van Gogh textual menu has been subdivided into four main categories, these are:- man-made objects, natural objects, people and buildings. On initially entering the textual menu subsystem the menu displayed contains four active buttons. These are shown in figure A4.7.

FIGURE A4.7



INITIAL VAN GOGH TEXTUAL MENU

From this initial textual menu one of these categories can be selected enabling specific objects within the Van Gogh pictorial database to be located. For example one specific object within the people category under object category PEOPLE object name MAN and sub category TYPES, is a FISHERMAN, therefore selection of this choice from within the people menu directly displays a picture of a fisherman from the laserdisk onto the television screen.

A4.1.4 The Relational Database

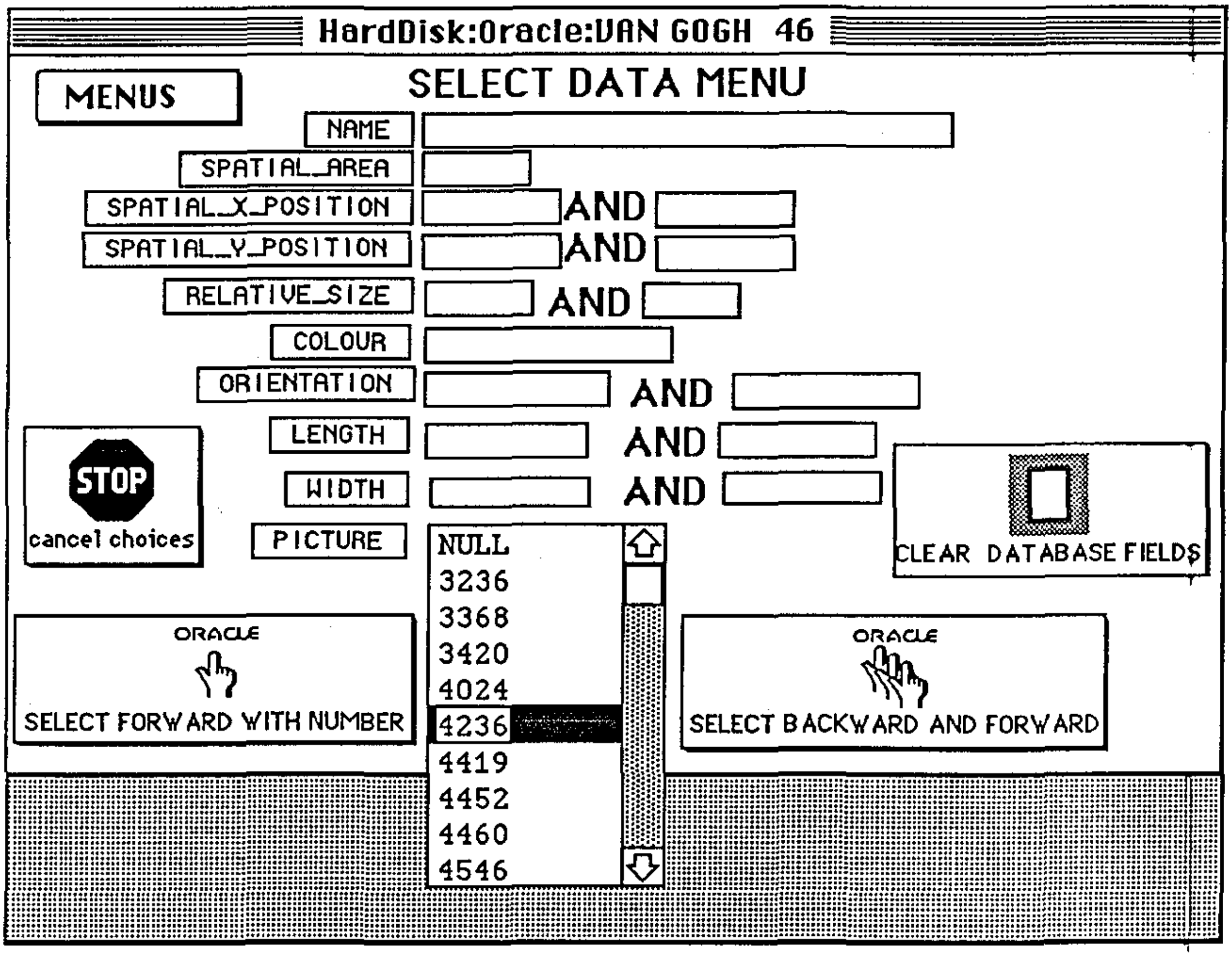
Oracle Corporation have recently developed a relational database that can be used within Apple's HyperCard. Software has been developed which allows users to directly enter values into fields from which records with matching field values can be displayed. At Loughborough University we have developed a relational database interface which acts as a series of structured query language (SQL) select statements. As discussed, values of physical attributes have been chosen to represent any picture in the pictorial database, these attributes (represented by fields in the Oracle database) are object name, spatial area, relative size, colour, orientation, length, width and laserdisk frame number containing the picture. The laserdisk frame number field within the interface allows the user to display all pictures which match the user's object query. Having manually encoded values of attributes into the relational database, user requested queries about these attributes can be directly entered into fields, from which matching information from the relational database can be found and the picture containing objects having these attributes displayed.

The size attribute encoded within the relational database was relative to the actual size of the screen which had been assigned the value of 10. Therefore the volume of the object's bounding box was relative to the overall size of the screen. For example a numerical value of 4 for an object's relative size meant that the bounding box of the object occupied 2/5ths (or 4/10ths) of the entire screen. The colour was the colour of the object being searched for. In many cases this was a large combination of colours and where this was so, this field was left blank, however there were many objects within the Van Gogh pictorial database which had a characteristic colour allowing users to search for them based upon the value of this attribute.

For example by entering the object name 'FLOWER' and the colour 'YELLOW' into the fields of the Oracle interface finds all pictures containing yellow flowers within the pictorial database and displays these pictures to the user. A facility has also been developed by which items can be selected from available values simply by clicking the mouse on the attribute field names present within the database application. By highlighting these field names the users can see at a glance all possible values within that field and they can then enter one of these into the SQL query by clicking the mouse on the appropriate one. This therefore acts as an alternative to searching through either the iconic or textual menus in order to find a specific example of the object or attribute required. This technique of selecting field names from the relational database interface is shown in figure A4.8.
In addition an interface was provided allowing the user to insert, delete and update Oracle database records. This interface is shown in figure A4.9. Attribute values could also be entered into the relational database fields by selecting or specifying information from either the visual or textual menus. For example within the

textual menu under the choice "MAN" the user is given two alternatives, either they can display all pictures containing men or alternatively attributes of men such as relative size or spatial location can be selected within the visual menus. On choosing the menu choice ATTRIBUTES OF MEN, the word MAN is automatically entered into the object name field within the SQL query and the user can then enter values of any other attributes about the picture which they know such as relative size, colour, spatial area or orientation.

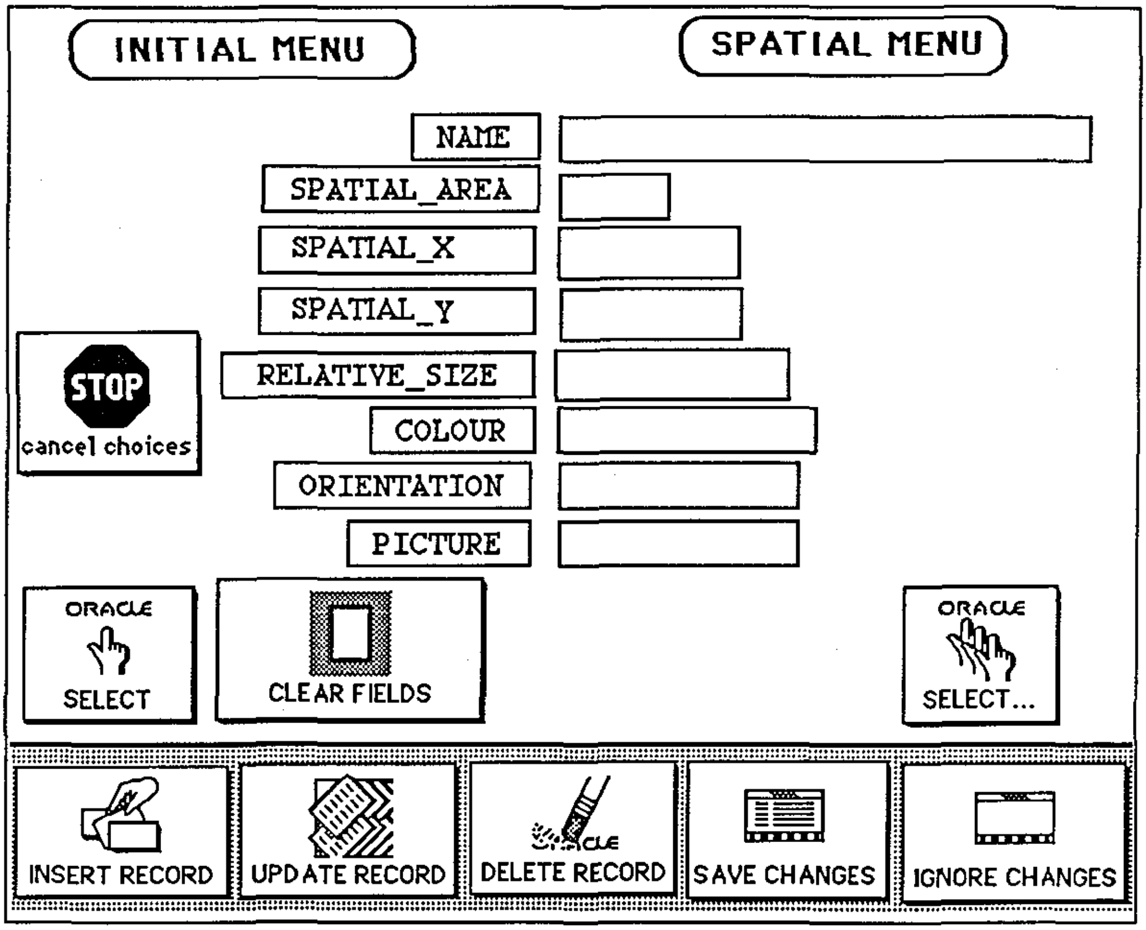
FIGURE A4.8



SELECTING A VALUE OF THE OBJECT PICTURE NUMBER FROM THE ORACLE RELATIONAL DATABASE INTERFACE

Having specified the known attribute values the user then has a choice of two Oracle select buttons enabling pictures containing these objects to be recalled from the laserdisk. The first select button containing an icon of a single hand allows the user to step forward through the matching pictures by clicking the mouse on the button while at the same time instructing the user how many records are left to view. The second select button containing an icon of multiple hands allows the user to move both backwards and forwards through the matching pictures so that direct comparisons between them can be made.

FIGURE A4.9



INTERFACE FOR DATA INSERTION, REMOVAL AND UPDATING OF INFORMATION WITHIN THE ORACLE RELATIONAL DATABASE

For both of the select buttons, at the same time that each picture is displayed, the corresponding record information from the relational database is also shown on the computer screen allowing the user a greater understanding of how the chosen objects and their attributes correspond to the pictures displayed. Once an object has been selected and displayed on the screen, the user can alter the values within the relational database table thereby recalling objects from within the pictorial database which bear a greater resemblance to the objects that they require.

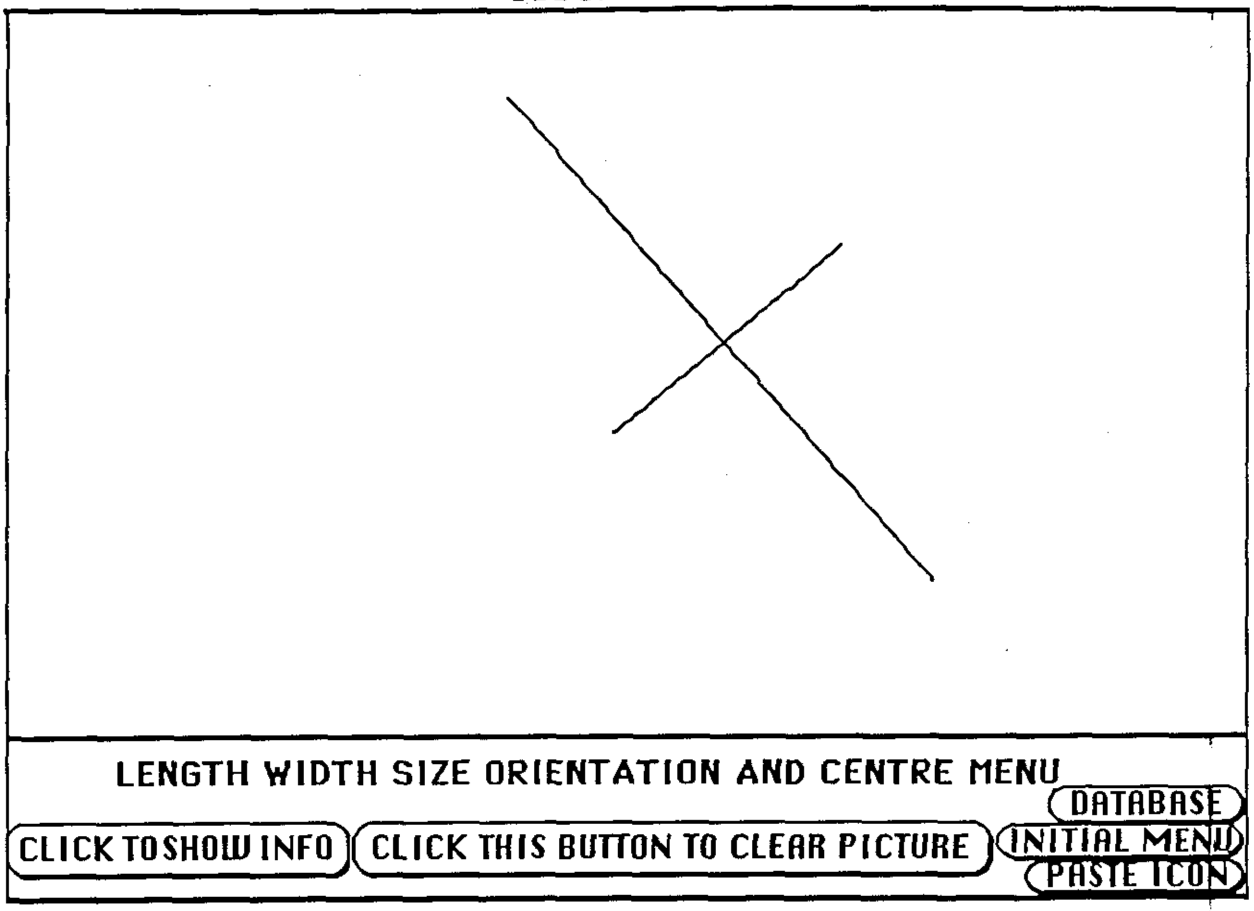
For example if the user first chooses the name 'TREE' for the object name field within the interface and clicks on one of the select buttons then all pictures of trees within the database will be found by the relational database management system and the user can step through each of these by continued pressing of the select button. However since the corresponding records are also displayed along with the pictures of trees, users can be more specific in their query by additionally selecting values of other attributes. For example if a SQL query which consists of only the word 'TREE' recalls over a hundred pictures of trees, users upon seeing the first picture of a tree of relative size 4, and colour 'GREEN' can then be more specific in their query. If these users require a tree twice the size of that displayed on the screen representing relative size 4 then a relative size of 8 is chosen from the relative size name field in addition to the object name 'TREE'. This query might only recall ten pictures of trees from the database. In order to locate a specific picture the users may then specify the value of another attribute such as an object orientation of 30 degrees to the horizontal. Thus the query would then be all GREEN TREES of relative size 8 and orientation 30 degrees to the horizontal, In this way the SQL query becomes progressively more detailed in order to find a specific picture from the laserdisk's pictorial database.

Therefore users can interactively select attributes of a picture, display matching pictures on the television screen and then alter the values of the attributes accordingly in order to locate a particular object within a picture that they require. When selecting a different object from the relational database the fields within the select query can be quickly cleared by clicking on a clear fields button which automatically clears all the fields on the screen. Having cleared all the fields, pressing one of the select buttons is equivalent to an ORACLE select query of * for all fields which selects all records stored within the database.

A4.1.5 Visual Depiction of Queries

An example of the visual depiction of a query is shown in figure A4.10 in which length, width, size, orientation, and absolute spatial location are depicted. The user first specifies the length and orientation of the object required by drawing a line. The second line drawn specifies the width. The size is calculated by multiplying the length by the width; while the absolute position of the object is obtained from the intersection of the two lines. This information is then entered into the Oracle relational database from which pictures containing objects having matching values of attributes are displayed.

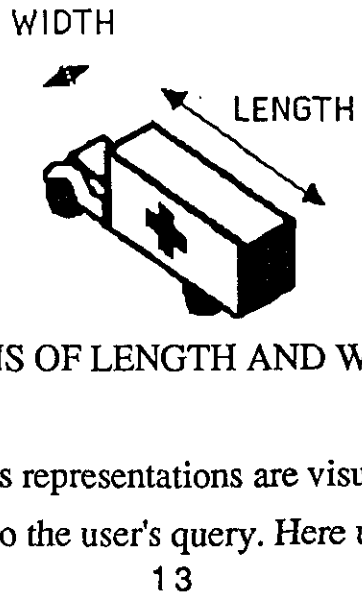
FIGURE A4.10



DEPICTION OF LENGTH WIDTH SIZE AND ORIENTATION

Thus for the object below the length and width are shown.

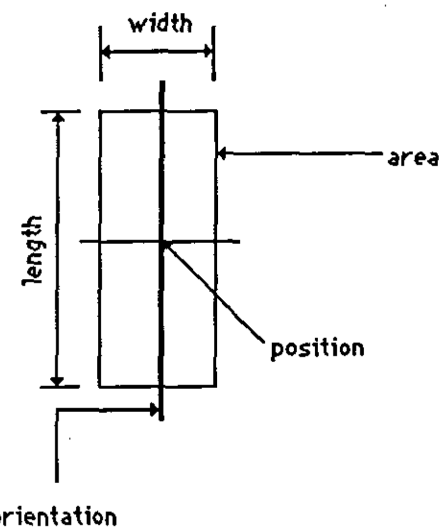
FIGURE A4.11



DEFINITIONS OF LENGTH AND WIDTH ATTRIBUTES

Under these circumstances representations are visually manipulated in order to more closely correspond to the user's query. Here users in effect 'sketch' their query by constructing a graphic depiction of the target picture rather than selecting from available choices from a menu. This allows users to indicate what they remember about an object rather than being restricted to a choice from a textual menu. Having manipulated this representation to closely correspond to their perception of the object within the picture required, this visual information can then be directly compared with information from the relational database and possible matching pictures displayed.

FIGURE A4.12



POSSIBLE DEPICTIVE SYMBOL

In this way visual and spatial attributes of objects in a picture can be specified depictively. Such a technique should allow users to describe all the information known about a required object within a picture. Again as with selection of positional information once the pictures have been displayed then choices can be refined since users can make visual comparisons between the actual pictures shown and the corresponding textual picture descriptions used within the fields of the relational database.

A4.1.6 <u>The Laserdisk and its Controller</u>

The use of a laserdisk as a picture source for an image database has been suggested by Canavesia and Marion (1989). Characteristics which distinguish a laserdisk from videotape are that each frame on the laserdisk had a digital frame number which could be randomly accessed and displayed as a frozen frame. This single frame capacity made it possible to store 54,000 individual images on one side of a disk. Thus a laserdisk is much more than a surrogate for tape, it can also be used as a storage mechanism for images in computational applications. The laserdisk player used within our system was a Phillips VP835 player connected to an Apple Macintosh IIx computer and an Imagewriter printer. The laserdisk player's RS232 port was connected to the modem port of the Macintosh. The player was also connected via an aerial lead to a 14 inch teletext colour television so that pictures from the player could be displayed onto the television. The volume of the sound on the laserdisk player was controlled by altering the control on the television, however the appropriate audio track was selected from an on-screen interactive laserdisk controller program on the computer. The hardware configuration used within our system is shown in figure A4.13.

Commands to the laserdisk player could either be entered from the computer or via a hand held remote control device supplied with the player. Two on-screen laserdisk controllers have been developed using Apple HyperCard, both allowing the user direct interactive control over the laserdisk. One controller operates simply by movement of the mouse over its pad, while the other requires that laserdisk player functions are selected by clicking the mouse on the appropriate area of the computer screen. The basic laserdisk controller functions were contained on a disk supplied by the Voyager Company of America.

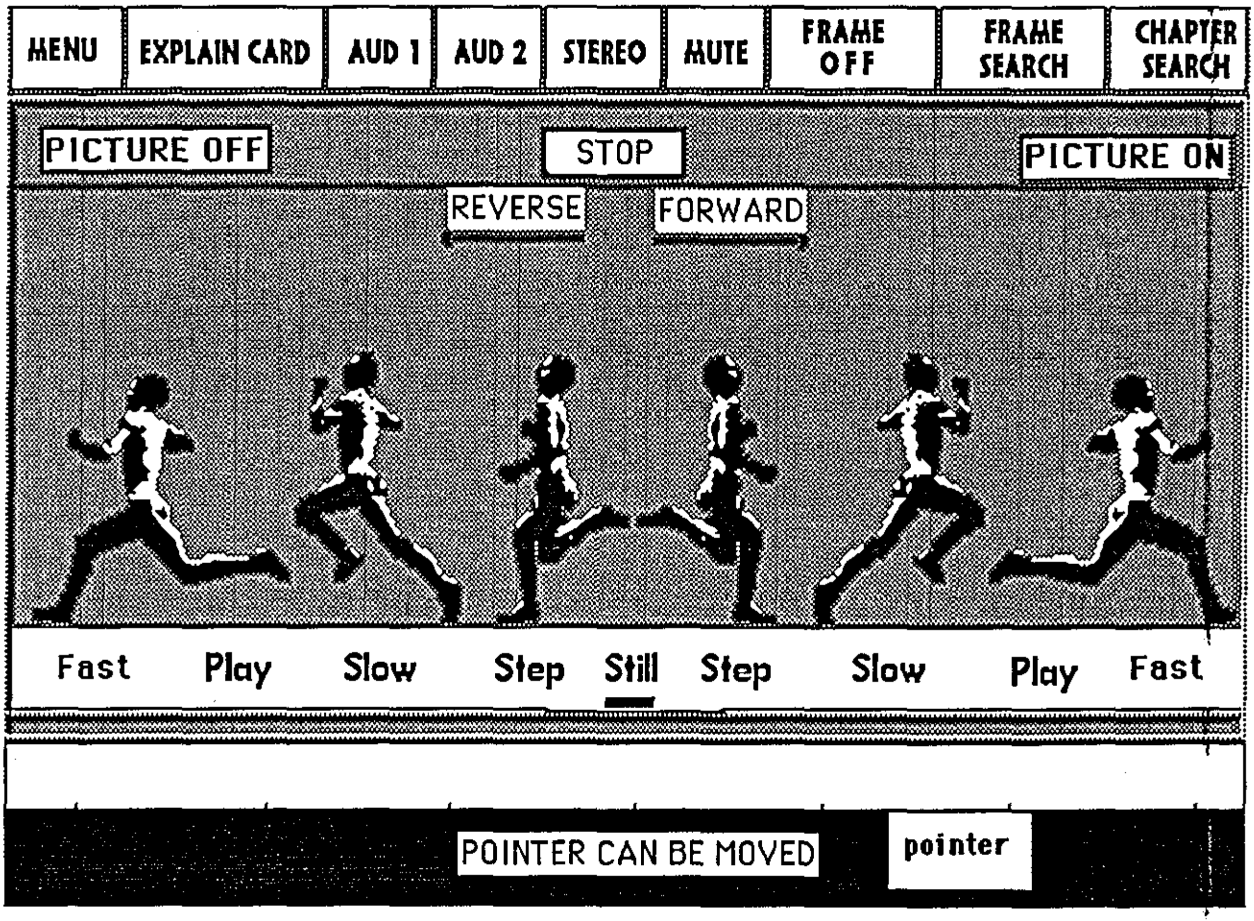
The first on-screen controller shown in figure A4.14 operates by moving the mouse within the area of the runners shown. On entering the areas the laserdisk automatically either speeds up, slows down or remains still at a frame without the necessity of clicking the mouse. Thus the user can scan through areas of the laserdisk either forwards or backwards both quickly and slowly and having located a frame of interest can then 'freeze frame' at that particular picture, the pointer shown below can also be dragged to a new frame number position within the laserdisk thereby allowing the user to move quickly to different areas within the pictorial database. Fast scanning in either a backward or forward direction could be achieved by holding the mouse down in positions in front of or behind the pointer.

FIG A4.13

HARDWARE CONFIGURATION FOR THE VAN GOGH PICTORIAL DATABASE

The various buttons at the top of the screen allow yet further control over the laserdisk. One of two audio channels can be selected giving different commentaries on the pictures within the disk. Sounds can also be muted or run in stereo in which both sound tracks present on the laserdisk are played simultaneously. The frame off button switches the frame display on the television off, this button then automatically changes is display to frame on, enabling the user to switch the frame display back on. The frame search button allows the user to specify a frame number within the disk after which that frame is displayed in still mode. Chapter search allows the user to choose a chapter within the laserdisk, while the picture on and off button switches the television picture on and off.

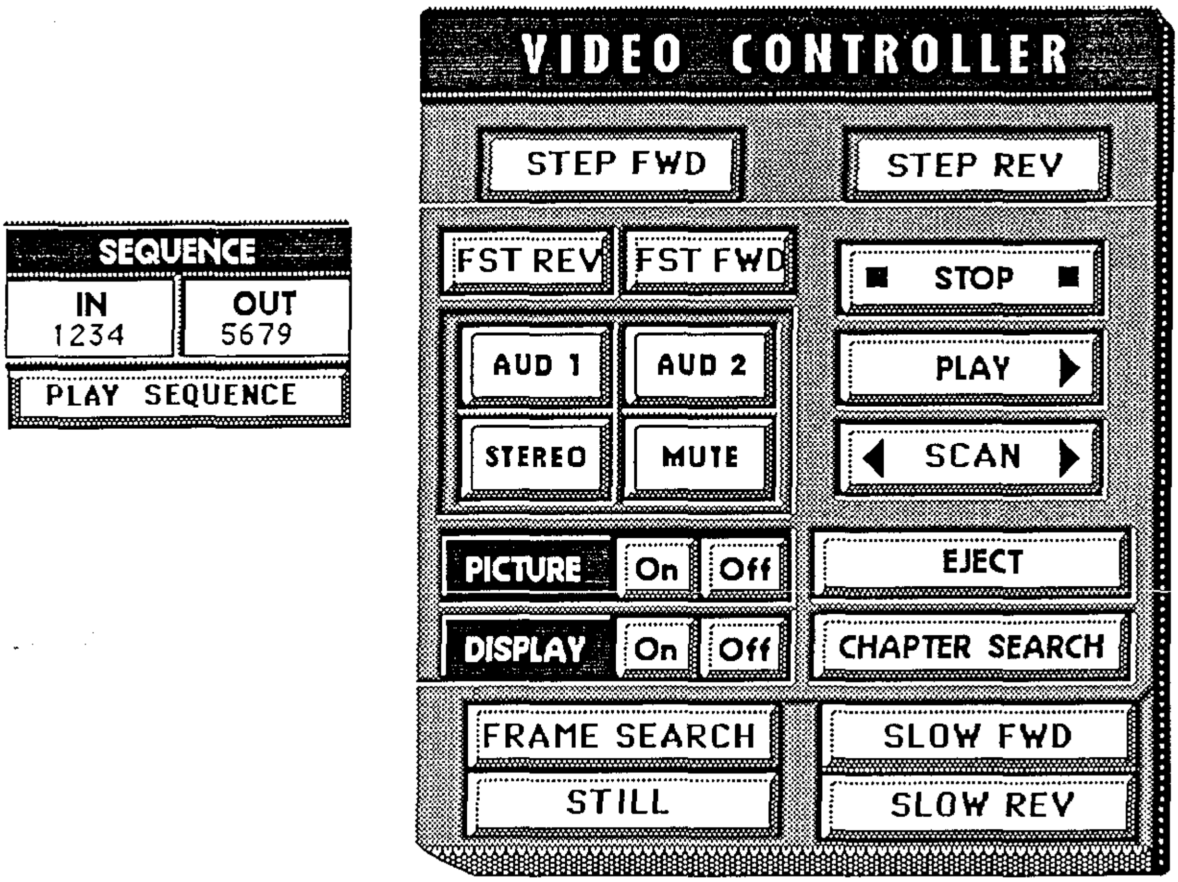
FIGURE A4.14



FIRST LASERDISK CONTROLLER

The second on-screen laserdisk controller has been fashioned in the shape of a hand held device. Here, in order to execute the relevant instructions to the laserdisk , the appropriate button within the on-screen controller is clicked. A facility has also been designed which allows the user to run a sequence on the laserdisk by entering the initial and final frames required. As with the first controller the laserdisk can be quickly scanned, played slowly in both the forward and reverse directions, stepped frame by frame in both forward and reverse directions, switched between audio channels, have both the screen display and frame number switched both on and off as well as the disk ejected. Overall this combination of on-screen commands for the two controllers allows complete control of the Van Gogh laserdisk provided.

FIGURE A4.15



SECOND LASERDISK CONTROLLER

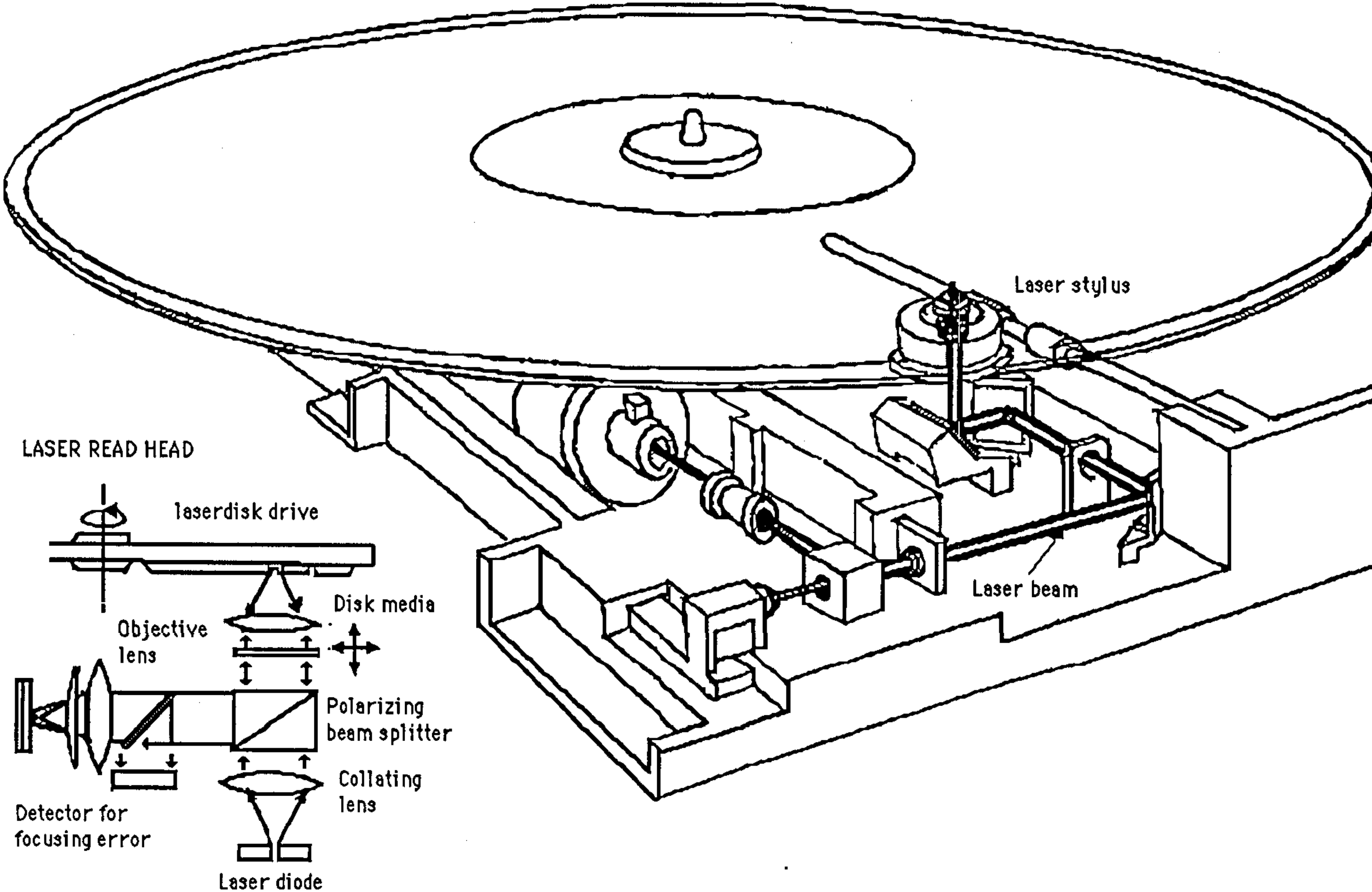


FIGURE A4.16
LASERDISK PLAYER

PAPER TO BE PRESENTED AT INTERACT (AUGUST 1990)

Using depictive queries to search pictorial databases

S. Charles and S.A.R. Scrivener

LUTCHI Research Centre
University of Technology
Loughborough

This paper argues that pictorial databases are becoming, and will continue to be, important in information systems because pictures can be used to depict information which is difficult to describe or perhaps incomprehensible in non-pictorial form. It follows that when searching for a picture the user might find it easier to depict the query by means of a picture. This paper describes a method for searching pictorial databases where the user essentially constructs a sketch (which combines depiction and description) of the target picture.

Introduction

In the past databases have consisted purely of text. However in recent years techniques for the storage, retrieval and manipulation of pictorial information within computers have become possible, due largely to the increase in resolution and processing power of the latest range of image display hardware. As a consequence it has become practical to construct databases that store large quantities of pictures. Chang [1] defines a pictorial database as an integrated collection of shareable pictorial data encoded in various formats to provide easy access by a large number of users. However, given technical feasibility, why should people wish to construct and use pictorial databases?

In some instances the answer to this question seems obvious: because the pictures exist. For example in remote sensing applications large quantities of digital image data are being constantly generated by sensing devices and surely these must be stored, ready for the user who wishes to analyse them. Another example of this kind that we can envisage is an art gallery, such as the Tate, that might wish to keep visual records of its collection. In this case a pictorial database is a necessity, is it not? Well in the case of remote sensing, if we were able to extract and describe all the relevant information in the remotely sensed image surely we could discard the picture. We may not be sure that we could do this, and if we could it might prove impossibly expensive; but at least in principle the proposition is logically reasonable. Perhaps we could treat the Tate's Picassos in the same fashion. To some this might seem to be the best thing to do with them; others, whilst perhaps being prepared to contemplate the idea of describing remotely sensed images will now balk at the suggestion. For them, at least, there is some fundamental difference between words and pictures; between description and depiction.

The difference between description and depiction has been the subject of considerable debate in discussions about representations Paivio, [2]; Palmer, [3] &[4]. Fish and Scrivener [5] summarise some of the suggested differences between descriptive and depictive representations. Descriptive representations involve sign systems, such as language, which have arbitrary learned rules of interpretation linking the sign system to the represented objects or concepts. Descriptions are useful for representing classes and properties of things. In addition, descriptive representations allow us to separate important from unimportant information. For example, specifying the relationship "on" without specifying position, or specifying the type of object without specifying its size or colour, say. The information in a description is extrinsic, meaning that it only exists by being associated with externally defined rules of interpretation.

In contrast, a depictive representation, sometimes termed analog, is not dependent on externally defined rules of interpretation because it causes visual experience which is similar to that associated with the object, or scene, or event represented. The colour of a cat may be described by the word "black" or depicted by spatially extended paint generating a similar colour sensation to the represented cat. Depictions represent spatial structure in a two or three dimensional spatial medium in which there are correspondences between spatial position in the medium and spatial position in the thing represented. Much of the information in a depiction is intrinsic, meaning it is not represented explicitly but can be extracted by inspection. Depictions are commonly associated with specific modes of perception. Apart from being necessary to represent detailed concrete spatial information, visual depictions facilitate the search for information not easily represented descriptively, or not easy to find because it is not represented explicitly. Perhaps, then, it is the depictive power of Picasso's painting that causes us to feel there is no descriptive substitute for his work.

It seems reasonable to suggest that often a picture is as much a thing of depiction as description. It is the potential of the picture as a medium of depictive representation which, in our view, provides the most compelling reason for their inclusion in information systems. Descriptive representations are amodal, they are not specific to, and as such do not benefit from the specific processing capabilities of a mode of perception. On the other hand visual depictions rely for their effectiveness on the processing power of visual perception and must be seen, hence pictures containing a depictive component must be available to the user. We shall argue here that "easy access" to pictorial databases implies an ability to search the database, when necessary, in terms of the depicted properties of the represented "world".

Pictorial database systems

Early pictorial databases consisted of both textual information and images. However, they were constructed by storing textual information separately from pictorial data, linked only by textual registrations of the images. Examples include medical databases storing X-ray images linked by textual registrations of patient name and patient number Assman, Venema and Hohne [6]. In such systems the user can only get to the picture by providing queries that match simple descriptions of the contents of the picture. Furthermore, it was not possible to search on depicted properties of the patient, such as the spatial arrangement of the heart and lungs, say.

Gradually systems using computerised digital analysis techniques were introduced in an attempt to automatically analyse the actual images within pictorial databases thereby converting from depiction to description. These systems have been further developed by the inclusion of textual picture query languages such as GRAIN Chang, Reuss and McCormick [7] and IMAID Chang and Fu [8] which enable users to construct their own queries about the information depicted within the images. IMAID for example is an integrated relational database system interfaced with an image analysis system. By using pattern recognition and image processing manipulation functions, symbolic descriptions of depicted structures can be extracted from images and stored in relational form. User queries about pictures can be manipulated through the relational database and pictures matching these queries displayed. In this way the need to process vast amounts of image data at query time is eliminated. If the user's requirements can be expressed in terms of the extracted descriptions, then there is no need to retrieve and process the actual pictures. If on the other hand the stored information is not sufficient, all pictures satisfying selection criteria can be retrieved and processed until the required precision is obtained. Such systems, which integrate conventional and picture query languages Chang and Fu [8], are flexible tools for analysing the contents of a pictorial database. However, such systems suffer from a number of limitations. The processes for extracting descriptions are highly application specific. For example, we would not expect the process that extracts the descriptor "highway" from a satellite picture to be very successful when presented with a picture of a highway taken from a land vehicle. Also, there is difficulty in extracting information about spatial arrangement depicted in the picture

The relational model used within such systems has been amongst the most popular techniques by which to analyse information within images. However, the use of relational calculus for manipulating locational data has been shown to have severe limitations Peuquet [9]. The basic set of operations of union, intersection and containment hold in a spatial sense, but this approach is derived purely from traditional mathematical concepts, and there is no ability to handle inexact, context-dependent relationships, set-oriented or otherwise, or of defining higher-order relationships on the basis of simpler in-built operators. Hence, even given that spatially depicted information could be translated into descriptions, the method of query provided to the user is unsuited to searching on these descriptions.

Meier and Ilg [10] have demonstrated that an extended relational database management system approach in which spatial relationships within a picture are directly encoded within a textual database is also severely limited. Such systems use a set of primitive textual relationships when storing spatial data Peuquet [9]. Here the designer views the spatial description of an entity as another attribute within the database. Various lists of "primitive" spatial relationships have been derived. These include spatial relationships such as "below", "left-of", "right-of", "above" etc. However, the difficulty of describing spatial relationships between entities in this way makes the use of such systems problematic. Here again, as for conventional text based relational systems, textual descriptions of relationships depicted in a picture are inadequate.

In current approaches to pictorial databases, then, there is a failure to handle effectively the problem of searching for a picture, or pictures, on the basis of depiction. The user has to describe the things of interest, and these descriptions are matched to descriptions derived from the pictures. However this is not fundamentally a machine problem. It may be that in the end, when all arguments are resolved, it will be demonstrated that everything that can be represented depictively can be represented descriptively. In such a case, it may prove possible to implement representational systems in the machine capable of fully describing what a picture depicts. The problem is that people find it difficult to use natural descriptive systems (such as language) to describe certain properties of the visual world or information depicted in pictorial representations. It is the nature of human perception and cognition that limits the usefulness of description and explains the potential of depiction. Consequently it is the nature of the pictorial query language that the user employs in searching the pictorial database that is at the heart of the problem.

Pictorial query and the sketch

As we have seen, previous pictorial query languages have been text-based; queries are entered by constructing sentence like statements. Of course, it was not the development of pictorial databases that first necessitated descriptions of the visual world. The very pictures stored in pictorial databases are often themselves examples of such representations (eg ordinance survey maps). More importantly, such representations are usually both descriptive and depictive. For example, on an ordinance survey map three towns will be identified by their names (description) and their relative distances from each other by the the relative distance between their locations on the map (depiction).

Typically visual design starts with a number of vague ideas which are clarified and developed with the aid of sketches. Fish and Scrivener [5] have argued that sketches are two-dimensional sign systems used, in general, to represent three-dimensional visual experience. They do this in two ways. For example, lines used in drawing have a variety of meanings which are (partly at least) culturally acquired. These descriptive sign systems are frequently supplemented with written notes. However, unlike purely descriptive sign systems such as writing, sketches are also depictive in the sense that they promote visual experience resembling that associated with the objects or scene represented. Sketching is a method by which a person can, with little material and some skill, produce representations of mental constructs. At any time the sketcher is free to use description and depiction as appropriate.

The sketch, then, provides an insight into how we might go about providing interfaces to pictorial database systems that allow the user to search for pictures (which are partly at least depictions) using a query method which is both depictive and descriptive. That is to say, the depictive component of the query might be provided graphically and descriptive component textually, or by the selection of other descriptive symbols, such as icons. In the following sections we describe a prototype pictorial database system that uses such a query method.

Picture properties

Before proceeding to describe our system it is important to consider exactly what spatial and visual properties of the target picture we are expecting the user to depict. Frequently a picture represents properties of a three-dimensional medium on a two-dimensional medium. One approach then is to allow the user to depict properties of a three-dimensional "world". Answering a query of this form would require machine processes capable of inferring properties of the three-dimensional "world" from two-dimensional depictions.

Although recognising the potential of such an approach we have not chosen to go in this direction. In the first place not all pictures are representations of a three-dimensional world, or the world "as seen". Many of the representational systems that humans have developed to communicate with other humans (eg circuit diagrams, flowcharts, data flow diagrams, architectural plans and elevations) represent non-visual systems, concepts and conceptual structures, or views of the world in which the third dimension is not relevant (eg architectural plans). In such instances extracting three-dimensional information from pictures is either meaningless or serves no useful purpose. In the second place, and with practicality in mind, we wished to provide some facility for querying by depiction quickly in order to investigate its value in practice, and mapping from two-dimensions to three-dimensions is by no means straightforward.

Instead we have chosen to provide methods that allow the user to enter two-dimensional properties depictively. The nineteenth century artist and critic Maurice Denis, in a statement that has often been used to justify the movement towards abstraction in painting, wrote that before being a horse or a battle a picture is a collection of shapes and colours arranged on a surface.

Similarly a picture can be described in terms of surface shapes, colours and arrangement. We have chosen to explore methods that allow the user to query by the depiction of pictorial properties.

Query features of the system

We can identify two extremes of query. At one extreme a query in terms of objects, attributes, and visual and spatial properties of a picture can be constructed purely descriptively. At the other extreme they can be constructed using depiction only. Between these two extremes lie queries that combine both description and depiction. In the following sections we provide examples of how the above query types are supported by our system.

Query by description

Descriptive tables are provided that allow the user to describe target pictures textually. This is achieved by selecting objects and attributes (including visual and spatial) by moving through a textual menu hierarchy. Words describing an object or its attributes can be typed directly into the appropriate field or, alternatively, the user can enter information into a field by selecting words from a mouse controlled pop-up menu. Thus, for example, clicking on the OBJECT name field would show all objects within the database, and clicking on the word LOOM puts the object name into the object field.

OBJECT	HAS-COMPONENTS	IS-PART-OF	TYPES
MAN-MADE OBJECTS			**JUG** **LOOM** **PIANO** **PICTURE** **PIPE** **PLATE**

Figure 1: Example of interface for constructing textual queries

Each textual menu within the hierarchy has several headings (see Figure 1).

Thus for the object flower; values in these headings might be

OBJECT	HAS-COMPONENTS	IS-PART-OF
FLOWER	STALK	FIELD
	ROOT	BUNCH

TYPES
SUNFLOWERS
IRISES
DAISIES

Selecting one of these values moves the user to a different menu until at the lowest level the actual images are located. Therefore the textual menus reflect the logical linking of data allowing the user to navigate through the database.

In practice the user has two initial choices when searching for a picture descriptively:

1) items can be selected directly from the relational database, or

2) objects can be selected by moving through the database structure via direct manipulation.

Using 1), on selecting an appropriate object name and choosing SELECT from the database management system, all records in the relational database which match the query will be displayed. Thus selecting 'fisherman' from the object name field will display all pictures within the database that contain fisherman. Other attributes can be entered if a more specific query is desired.

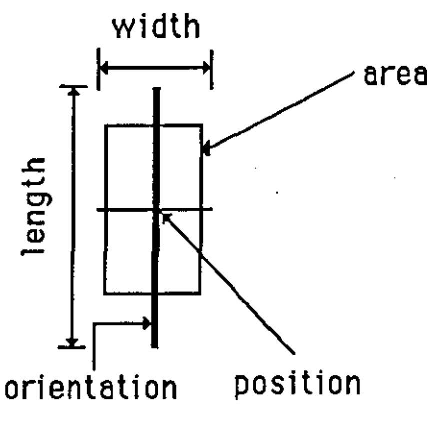
Using 2) the user moves through the textual database by directly manipulating the menus in order to locate a specific object after which attributes of that object can be selected. Thus for example moving through the hierarchical menu might involve selecting the following:

PEOPLE------>MAN--------->FISHERMAN

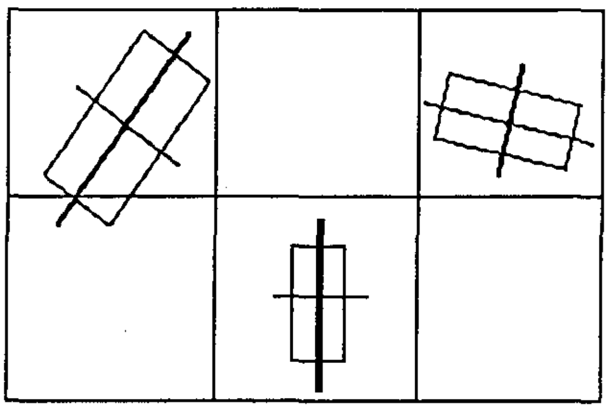
Having identified the object (in this case a fisherman), visual and spatial attributes (currently relative size, orientation, colour, length, width, and position) can be entered into the appropriate relational database query field. Thus at present the text menu structure is designed to assist the selection of an object name, and it is only after this is done that object attributes are entered by typing in values.

Query by depiction

Pure depictive techniques of picture query allow the user to construct a graphic depiction of the target picture. Shapes on a picture surface can be depicted by manipulating a depictor, Figure 2a. The rectangular part of the symbol depicts size, the bars depict length and width and their intersection the position of a shape. The thicker bar depicts the orientation of a shape and by default its length. Size can be specified by varying the rectangle; width and length by varying the bars; and orientation by rotating the entire symbol, Figure 2b. In this way visual and spatial attributes of shapes in a picture can be specified depictively. Currently, no facility exist for specifying colour



2a



2b

Figure 2: Depictive symbol and example of its use in pure depictive query

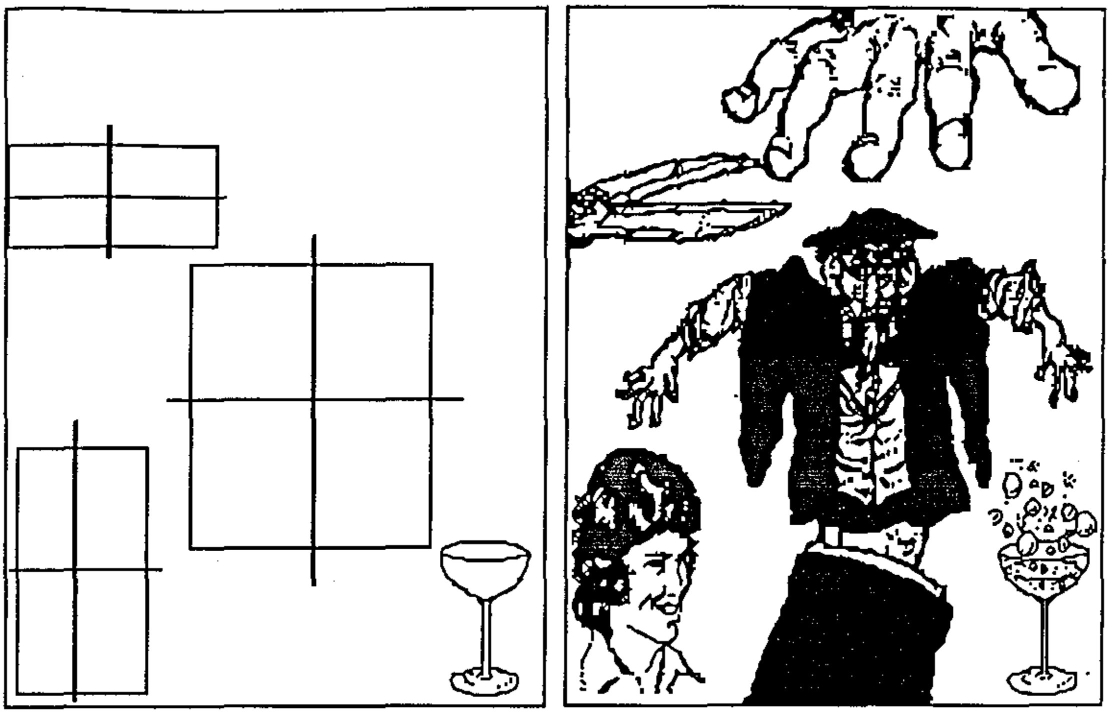
The method by which a query is constructed using purely depicted techniques is shown in figure 2b. First the user selects the entity depiction symbol and locates it in an area of the screen representing the picture surface (which can be user defined). Currently this field is divided into a number of sectors and when a symbol is placed its location is defined by the label of the section in which it falls. Many such symbols can be placed into the picture field and manipulated, thus allowing the user to depict visually complex queries.

In addition, icons depicting objects can be selected from menus and located in the picture field and subsequently manipulated. In this way combinational descriptive and depictive queries can be used to identify pictures from a pictorial database, since the icons have names attached to them and hence placing a wine glass icon in the picture surface area is equivalent to writing the words "wine glass". Figure 3a, which is a query for figure 3b, illustrates this idea. Here some objects have been identified by manipulation of the shape symbol. Also a representation has been selected from an iconic menu and placed in an appropriate position in the picture field. This type of query corresponds closely to a sketch in the sense that it combines both descriptive and depictive components.

Refining choices

One of the major advantages of such a system is that choices can be refined. When a query is satisfied the first picture in the set of "matching" images is displayed. The user can compare the picture to the specification in the table, and as a consequence refine the query.

This facility is probably more applicable in the descriptive case rather than the depictive, since in the depictive case the pattern that the user constructs should match closely the patterning in the pictures retrieved from the database. However, where a description of a picture is used the user will not know how these descriptions will appear (in the sense of predicting matching pictures). On seeing a "matching" picture displayed the user can enter more accurate quantities based on a comparison between the displayed picture and the mental image of the target shape.



3a 3b

Figure 3: Querying by depiction and description

For example, on selecting an object name "loom" descriptively (either textually or iconically) the system retrieves twenty pictures containing a loom. Seeing that the picture of the loom on the screen is of size six (say) the user may conclude "well I want a loom much smaller than that" and consequently enters two as the new loom size. Other attributes can be altered in this way, so refining the query on the basis of what is seen on the screen.

System configuration

The system is implemented on an Apple Macintosh IIx connected to a Philips VP835 laserdisk. Information from the laserdisk is displayed on a 14 inch colour television. The application runs on HyperCard version 1.2 and the relational database (built using ORACLE, version 1.1 for Macintosh) stores descriptions of paintings by the nineteenth century post-impressionist artist Van Gogh, recorded on the laserdisk titled "Vincent Van Gogh (a portrait in two parts)", published by North American Philips Corporation.

Entering picture descriptions into the database

So far we have described how the user can query by depiction but we have not explained how the descriptions of the pictures against which queries are matched are entered into the database. For the Van Gogh database this was done manually. All pictures were inspected in order to identify objects, and quantify, by visual judgment, the location and attributes of shapes. Clearly, this was a time consuming process but was adequate for our purposes. In the future we will explore a number of ways of simplifying this activity. As we have already mentioned, a sketch provides a way of representing a picture. It can be used to represent a picture in mind (for search) or a visible picture. Initially, we propose to modify the system so that a picture can be entered into the database by constructing a sketch. In this later case the sketch will be constructed over a displayed image of the picture that is to be acquired.

We will also investigate the use of image analysis techniques Scrivener and Schappo [11]. This will allow greater drawing freedom using a painting system. Here a sketch will be painted and the sketch processed automatically to derive shapes and attributes of shapes. Methods for doing this have already been developed Woodcock et. al [12], what remains to be done is to implement them for the pictorial database application. It is possible to get the descriptions directly from the picture using these techniques, but we foresee difficulties with this and prefer to follow the more practical path described above.

Conclusions

We have argued that information technology systems are capable, in general, of storing at a low cost large volumes of pictorial data. We have suggested that because of the depictive power of pictures we should anticipate a growth in the use of pictures and hence an increased reliance on pictorial databases. However, pictorial databases will only be useful to the extent to which they allow the flexible and rapid search.

We have drawn a distinction between depictive and descriptive representations and have argued that pictures are important because they, in part at least, depict and hence allow the viewer to take up information more efficiently because of the way in which they address visual perception and cognition. It follows from this that when searching for a target picture the user might find it easier to depict the picture (since a picture is itself a visual and spatial thing and pictures are useful for representing such things) rather than, or as well as, describing it. Earlier methods for accessing pictorial databases make no provision for depictive search. In this paper, we have described a system that provides methods that include search by depiction. Essentially the user creates a "sketch" of the target picture from which the system extracts descriptions that are matched to descriptions of the pictures in the database.

Currently a shape and its gross attributes, including position, can be specified depictively. As a consequence the system is application independent. In the future we will attempt to maintain this application independence whilst investigating how other properties of a picture might be communicated depictively, including for example shape, and relationships between shapes (eg adjacency), and how to improve the ease with which descriptive-depictive queries can be constructed and manipulated.

[1] Chang, S.K., Pictorial Information Systems. Computer. November 1981. pp.10-11.

[2] Pavio, A., Imagery and Verbal Processes (Holt, New York,1971).

[3] Palmer, S.E., Visual Perception and World Knowledge. Notes on a Model of a Sensory-Cognitive Interaction , in: Norman,D.A. and Rumelhart,D.E. et al (eds.), Explorations in Cognition (Freeman, San Francisco, 1975).

[4] Palmer, S.E., Hierarchical Studies in Perceptual Representation, Cognitive Psychology, No 9, (1977) pp.441-474.

[5] Fish, J. and Scrivener, S.A.R.S., Amplifying The Mind's Eye; Sketching And Cognition. LEONARDO. (to appear) (1990).

[6] Assman, K. Venema, R. and Hohne, K.H., The ISQL Language: A Software Tool for the Development of Pictorial Information Systems in Medicine, in: Chang,S.K. Ichikawa,T. and Ligomenides,P.A.(eds.), Visual Languages (Plenum Press 1986). pp.261-284.

[7] Chang, S.K. Reuss, J. and McCormick,B.H., Design Considerations Of A Pictorial Database System. Policy Analysis and Information Systems. vol 1, No 2, (1978) pp.49-70.

[8] Chang, N.S. and Fu, K.S., Picture Query Languages for Pictorial Data-Base Systems. Computer. November 1981. pp.23-33.

[9] Peuquet, D.J., Towards The Definition And Use Of Complex Spatial Relationships. Proceedings of the Third International Symposium on Spatial data Handling. August 17-19. (1988) pp. 211-223.

[10] Meier, A. and Ilg, M., Consistent Operation on a Spatial Data Structure. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol 8. no 4 (1986) pp.532-538.

[11] Scrivener, S.A.R. and Schappo, A., Perceptual Approach to Picture Interpretation'. Knowledge Based Systems. Vol 1. No 2 (March 1988) pp 105-113.

[12] Woodcock, A. Plant, T. Schappo, A. and Scrivener, S.A.R., A Psychological Oriented Approach to Picture Interpretation. Poster Paper. HCI 88, Manchester. (1988) pp. 1-12