Searching Pictorial Databases by Means of Depictions

"Stephen Scrivener and ""Stephen Charles

LUTCHI Research Centre
University of Technology
Loughborough

" artist , computer scientist
""computer scientist

This paper argues that, in general, information technology systems are capable of storing at low cost large volumes of pictorial data and hence we should anticipate an increased reliance on pictorial databases. However, pictorial databases will only be useful to the extent to which that allow flexible and rapid search. A distinction between depictive and descriptive representations is made and it is argued that visual depictions, such as drawings, are useful for representing visual-spatial properties of pictures. Earlier methods for accessing pictorial databases make no provision for depictive search. In this paper, a system that provides methods that include search by depiction is described. 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.

INTRODUCTION

The range of domains in which computers are applied has increased quite significantly during the last decade. This is partly due to increased availability as a result of low cost miniaturisation. At least as important, however, in the spread of computers is the increased range of information representation formats that can be supported. In the early days information had to be represented using numerical codes, later words, and more recently pictures. Today even personal computers can handle the creation, manipulation, presentation and storage of pictures, making the use of computers in art and design a commonplace occurrence. This increased use of computers by people who routinely generate and use pictures has led to the need for pictorial databases that store large quantities of pictures. Clearly there is a role for pictorial databases in many domains. However, in this article we only consider issues relating to databases of paintings, although we hope to show that the ideas discussed here are of more general interest.

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. With large volumes of pictorial data "easy access" becomes a crucial matter. In such cases it is usually not practical to scan all pictures in order to locate the one or more pictures desired. Instead, to make search practicable and timely the pictorial database is organised so that various selection strategies can be employed by the user to locate a smaller set of pictures for viewing. The goal of a database designer is to provide a search strategy that produces a picture set containing only pictures, and omitting none, that match the user's target picture (set), or is empty in the event of there being no matches in the database. We shall argue here that "easy access" has a lot to do with the 'language' available to the user for communicating the target picture, or picture set, definition to the computer.

SEARCHING ON VISUAL-SPATIAL PICTURE PROPERTIES

The two-dimensional visual-spatial properties of the picture surface are important to artists and art historians. Indeed in the development of twentieth century painting notions of picture surface have played a crucial role. In 1890 the painter-aesthetician Maurice Denis wrote, " Remember that before being a war-horse, a nude or some story or other, a picture is essentially a flat surface covered with coloured pigments arranged in a certain order [2]. In this statement Denis defined the essential principle of modern painting. "The moderns discarded any dependence on subject or nature and made it their business to create 'pure painting'. Even when they wanted to translate some definite emotion, it was the form itself that had to be expressive, and not the evocation of some touching theme" [2]. However, this preoccupation with the structure and organisation of the picture surface is not restricted to twentieth century painting. Throughout the history of painting the organisation of represented objects and scenes has been partly constrained by the need to comply with a required surface order. It is not unreasonable to postulate that a user of a database of paintings might wish to search for pictures having particular two-dimensional visual-spatial properties. To what extent, then, is visual-spatial search supported in existing pictorial database techniques?

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 [3]. In such systems the user can only get to the picture by providing queries that match simple descriptions of the picture content.

Gradually systems using computerised digital analysis techniques were introduced in an attempt to automatically analyse the actual images within pictorial databases. These systems have been further developed by the inclusion of textual picture query languages such as GRAIN [4] and IMAID [5] 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 imagery data at query time is eliminated. If a 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. Systems, which integrate conventional and picture query languages [5], 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 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 location data has been shown to have severe limitations [6]. 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.

Meier and Ilg [7] 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 [6]. 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.

In current approaches to pictorial databases, then, there is a failure to handle effectively the problem of searching for a picture in terms of its visual-spatial structure. Where facilities are provided for querying on visual spatial properties these are restricted to simple linguistic terms. The user has to describe the things of interest, and these descriptions are matched to descriptions derived from the pictures. The problem is that humans' find it difficult to use natural descriptive systems (such as language) to adequately describe the visual-spatial structure of a picture. To provide "easy access" of a visual-spatial kind to a pictorial database we need to consider ways, other than language, for representing the target picture.

DESCRIPTIVE AND DEPICTIVE VISUAL REPRESENTATIONS

In recent literature dealing with the nature of representations there has been much discussion of the difference between description and depiction [8,9,10]. Fish and Scrivener [11] summarise some of the suggested differences between descriptive and depictive representations. Descriptive representations involve signs 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. As we have seen, pictorial database techniques rely heavily on the use of descriptive representations (eg. textual queries).

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. Visual 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.

REPRESENTING VISUAL-SPATIAL STRUCTURE

For centuries artists have used drawing as a way of depicting the visual world. More recently, in design, drawings have been used extensively to represent imaginary worlds. Drawings are depictive in the sense that they promote visual experience resembling that associated with the objects or scene depicted. For example, there will be visual correspondences between drawn contours of a copy of a picture and reflectance contours of the original picture in terms of curvatures, relative lengths and distances. In general, these properties of the picture are not represented explicitly in the drawing. Instead they are implicit in the sense that they are reflected amongst the array of structures constructed by visual perception when presented with the drawing. For this reason drawing looses some of the precision of a descriptive representation, since each viewer may take different accidental visual structures as representational. On the positive side drawing gains in the sheer volume of available visual-spatial structure in apparently immediate percepts that can be used to represent.

A simple task for the reader will perhaps illustrate this point. Take a sheet of paper and draw the outline of a circle on it. Now draw a second circle to the right of the first, and finally a third circle between and above the two already drawn. Immediately we see more in the drawing than is explicitly stated in the above instructions. We see a triangular arrangement of circular shapes located in particular relation to the limits of the paper, in a particular orientation, and having a particular shape. If one only wanted to represent the information explicitly stated in the instructions the depiction might be misleading as other perceived properties, such as those cited above, may not be properties of the represented world. Also the properties that come to the attention of one reader when viewing his drawing might be quite different to those noticed by another in his, even if their respective drawings were exactly the same. In contrast, the incompleteness of the above instructions as a representation of the visual-spatial structure of a picture can now be seen by comparing the drawing produced from this description to the picture described (Fig. 1). It would be interesting to collate all readers' drawings in order to compare the difference in appearance between them. One assumes that there would be considerable variation between them. To reduce this variation and to produce closer copies many more instructions would be required for even this simple picture, hence increasing the effort required in constructing the description. However, it is not difficult to see that constructing a drawing of it (Fig. 1) would be a more straightforward task for the communicator, even one with little drawing skill, and more likely to replicate its visual-spatial appearance. Hence we posit that depiction, in the form of drawing, should be considered as an alternative means of representing the visual-spatial structure of a target picture, or pictures, required by the user of a pictorial database. Put another way, we propose that the user queries the pictorial database by constructing a drawing of the target picture, or pictures. However, what evidence is there to suggest that humans would be able to envision such depictions?

DRAWING FROM IMAGINATION AND THE SKETCH

Imagine a user about to search a pictorial database. In front of him or her is a photograph of a picture the user hopes to find in the database. Here envisioning the query should not be too difficult as a version of the target picture is available to perception. However, what if there is no such external copy of the target picture. In this case the user must construct the query from memory. The question arises as to whether mental representations are likely to support the construction of such queries. Most of us can answer this question from experience. Imagine your favourite picture and there is a sense of a picture in the mind; a mental image of the picture. A mental image that whilst less stable and vaguer than the percept of the picture in some sense resembles the picture. There is now considerable evidence [12,13,14] that the type of mental images that resemble percepts are 'quasi-pictorial' spatially depictive maps in which size, shape, and relative distance of a visual element are implicit in the position of a matrix of neural elements. These internal visual depictions seem to share many of the properties of external visual depictions, such as drawings. Given that mental images resemble percepts and that constructing drawings from percepts is a relatively straightforward matter it would seem that the idea of doing the same from mental images is not beyond the bounds of reason. Indeed, we have clear evidence for this human ability in the drawings created by designers to represent imaginary, or non-existent objects.

Typically such drawings are depictive in the sense that we have described. However, they are also descriptive in that often the elements of the drawing only partly resemble the thing represented, for example the window symbol in an architectural sketch. These elements of the representational scheme are partly descriptive and partly depictive, and have explicit rules of interpretation which must be learnt by the interpreter. In addition the drawn parts of the sketch are often supplemented by purely descriptive representations such as written notes. Hence, in general, sketches used to depict imaginary objects are partly depictive and partly descriptive representations. Clearly, we would not wish to suggest that depiction replaces description as a means of searching pictorial databases, rather that depictions can provide a useful adjunct to descriptive methods. The sketch, then, provides an insight into how we might go about providing interfaces to pictorial database systems that allow a user to search for pictures 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 the 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 provides such a query method for visual-spatial search. 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. For reasons of convenience, however, the images used below are not taken from this laserdisk.

QUERY BY DESCRIPTION

We can identify two possible extremes of query. At one extreme a query in terms visual-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 purely descriptive and descriptive-depictive queries are supported by our system.

Descriptive tables are provided that allow a 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 (Fig. 2), and clicking on the word LOOM puts the object name into the object field.

Each textual menu within the hierarchy has several headings. Thus for the object flower; values in these headings might be

OBJECT	HAS-COMPONENTS	IS-PART-OF	TYPES
FLOWER	STALK	FIELD	SUNFLOWERS
	ROOT	BUNCH	IRISIS
			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 a user to navigate through the database.

In practice a 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 a 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 DESCRIPTION-DEPICTION

Pure depiction is not supported by our system. Instead depictive-descriptive query is provided by means of graphic objects (Fig. 3a) that can be located and combined (Fig. 3b) in a display area representing the picture surface of the target picture. Visualised shapes of the target picture can be depicted by locating these graphical entities in this display area. As we can see (Fig. 3a), a thin line represents the width of a shape. Used on its own, its middle point represents the position of the shape. In a similar way a thick line represents the length and position of a shape, and in addition its orientation. When the width and length symbols are used together in the specification of a shape their intersection point marks the position of the shape and their product its size. The thick rectangle is used to represent tolerances on the position of a shape in the database that matches the query. Finally, the thin rectangle is used to represent size in the absence of either the width or length symbol. Default tolerances are associated with the length, width, position and size attributes.

Clearly, these symbols are partly descriptive and partly depictive. For example, there is no correspondence between thickness of line (in the representation) and width and length (in the represented world), and the user must learn to associate different thicknesses with the related attributes of length and width. In this sense, line thickness is used descriptively. On the other hand the metric distances defined by the length and width lines do share correspondences with the represented world. A 'width' line in the query having the property of being greater than another 'width' line in the same query depicts a picture in which the perceived width of the shape corresponding to the first line would be greater than that of the shape corresponding to the latter. Similarly, the orientation of the 'length' line depicts the orientation of a corresponding shape. Hence we would expect some perceived correspondence between the orientations of the representational symbol and a matching shape; they should appear perceptually similar.

In order to construct a query, first the user selects a drawing tool representing the attribute required and locates it in an area of the screen representing the picture surface (which can be user defined). Many such symbols can be placed into the picture field and manipulated, thus allowing the user to depict visually complex queries.

The use of these graphical entities is illustrated in a query (Fig.4) constructed with reference to a watercolour painting of Lyme Regis, by Scrivener (Fig. 5). Here, starting from the top left (comparing Fig. 4 with Fig. 5), the position, size, length, width and orientation of the bush is represented by combining the length and width markers, as are these attributes of the cliffs in the middle distance. Only the positions,lengths, and orientations of the buttress at the bottom left and breakwater in the middle right are defined, whereas the window group at the top left is specified in terms of size, and position tolerances. Finally, all of the visual-spatial attributes of a shape (ie length, width, orientation, size and position tolerances) that can the specified using the system are defined for the breakwater at the bottom right of the picture.

If this example (Fig.4) was used to query a pictorial database it would cause pictures consisting of shapes that matched those of the query to be retrieved irrespective of the content of the matching pictures. For example, a still life might be retrieved. In many instances this might be exactly what the user wants. In other instances one can imagine that the user might wish to retrieve pictures that consist of particular objects having particular visual-spatial characteristics. As we have seen above, this can be achieved by first selecting objects using the textual menus and then proceeding to define visual properties. An alternative approach to this is illustrated (Fig. 6) in which icons of objects are selected and located in the query. Essentially these icons are descriptions since they are equivalent to writing 'window' or 'fence'. In this way combined descriptive and depictive queries can be used to identify pictures from a pictorial database.

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 judgement, 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 [15]. 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 [16], 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.

CONCLUSION

We have argued that information technology systems are capable of storing at low cost large volumes of pictorial data and hence we should anticipate an increased reliance on pictorial databases. However, pictorial databases will only be useful to the extent to which that allow flexible and rapid search.

We have drawn a distinction between depictive and descriptive representations and have argued that visual depictions, such as drawings, are useful for representing visual-spatial properties of pictures. 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.

REFERENCES

1. S.K.Chang, "Pictorial Information Systems", *Computer*, November (1981) pp.10-11.

2. G. Bazin, *A Concise History of Art* (London: Thames and Hudson, 1958).

3. K. Assman, R.Venema, and K.H.Hohne, "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.

4. S.K. Chang, J. Reuss and B.H. McCormick, "Design Considerations Of A Pictorial Database System". *Policy Analysis and Information Systems,* **1**, No 2, (1978) pp.49-70.

5. N.S. Chang and K.S. Fu, "Picture Query Languages for Pictorial Data-Base Systems", *Computer*, November 1981, pp.23-33.

6. D.J. Peuquet, "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.

7. A. Meier and M. Ilg, "Consistent Operation on a Spatial Data Structure". *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, **8**, No 4, (1986) pp.532-538.

8. A. Pavio, *Imagery and Verbal Processes* (Holt, New York,1971).

9. S.E. Palmer, "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).

10 S.E. Palmer, "Hierarchical Studies in Perceptual Representation", *Cognitive Psychology*, **9**, (1977) pp.441-474.

11. J. Fish and S.A.R. Scrivener, "Amplifying the Minds Eye: Sketching and Cognition", *Leonardo*, 23, No.2. (1990) pp 117-126.

12 R.N. Shepard and L.A. Cooper, *Mental Images and Their Transformation* (Cambridge. MA: MIT Press, 1980).

13 S.M. Kosslyn, *Ghosts in the Mind's Machine* (New York: Norton 1983).

14 S.M. Kosslyn, *Image and Mind* (Cambridge MA: MIT Press 1980).

15. S.A.R. Scrivener and A.Schappo, "Perceptual Approach to Picture Interpretation", *Knowledge Based Systems*, **1**, No 2 (1988) pp 105-113.

16. A. Woodcock, T. Plant, A. Schappo and S.A.R. Scrivener, "A Psychological Oriented Approach to Picture Interpretation", *HCI 88*, (1988) pp. 1-12.

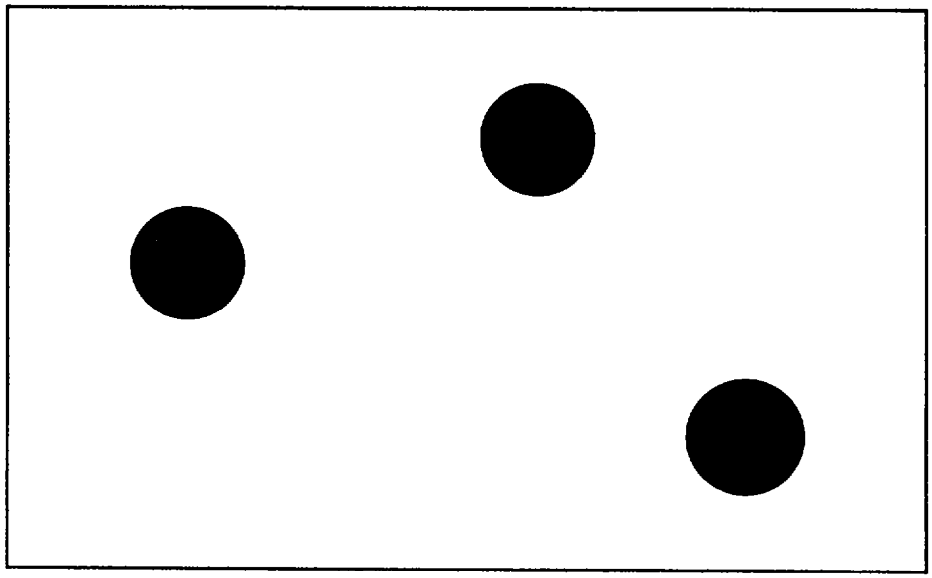

Fig. 1. This picture can be described as "a circle with another circle to the right of it, and a third above and between these two". However, the inadequacy of such a textual description as a representation of the picture's visual-spatial structure can be demonstrated by using it to produce a drawing without sight of the original picture described. A comparison of the drawing and the picture is likely to yield many differences. To reduce these differences many additional descriptive statements are required even for such a simple picture as this. In contrast, it is much easier to construct a representation of the visual-spatial structure in the picture by drawing.

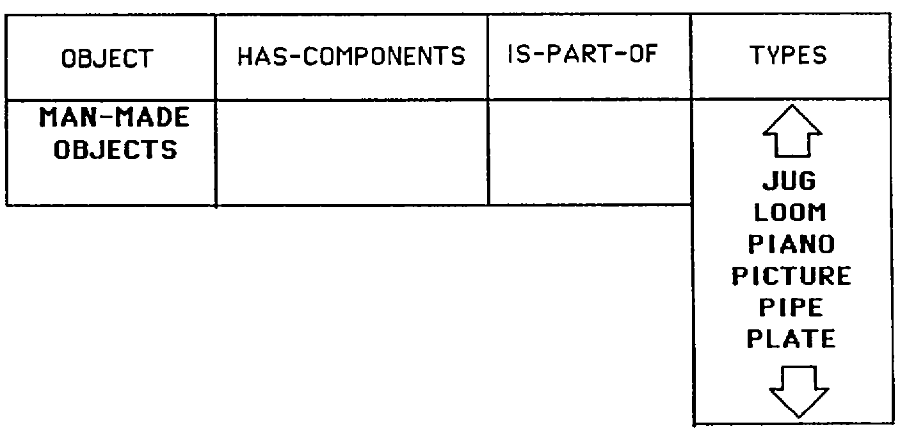

Fig. 2. Descriptive search. The system allows the user to issue queries by selecting objects via a menu system linked to object type and part-whole heirarchies. Here, having inserted "man-made objects" as the point of interest a listing of the next level down in the hierarchy can be scanned and an object type selected, if desired, for insertion into the 'object' field.

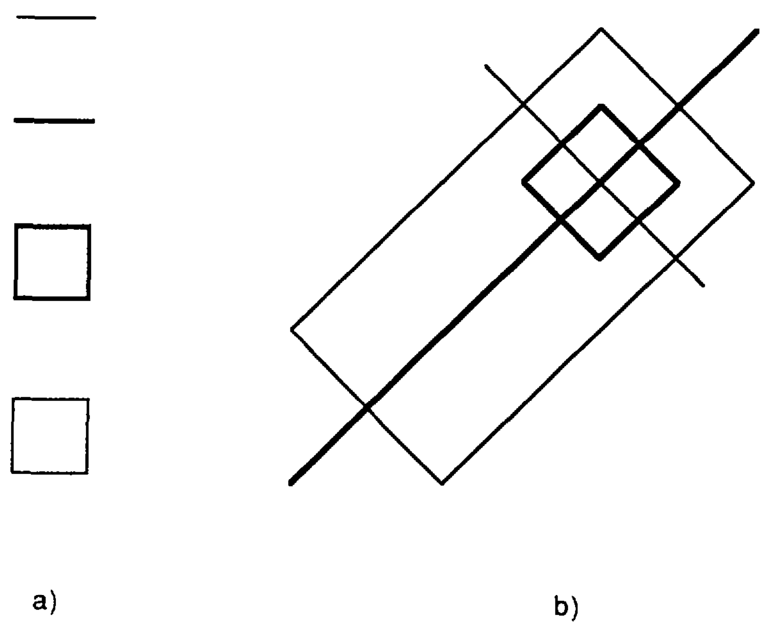

Fig. 3. Shape depiction entities and their combination. Starting from the top left of a) a thin line is used to represent shape width and position; a thick line length, position and orientation; a thick edged rectangle shape position tolerances, ie. the centre of gravity of a matching shape in the database can fall anywhere within this rectangle; and a thin edged rectangle the size of a shape. These shape depictors can be used individually or combined as in b).

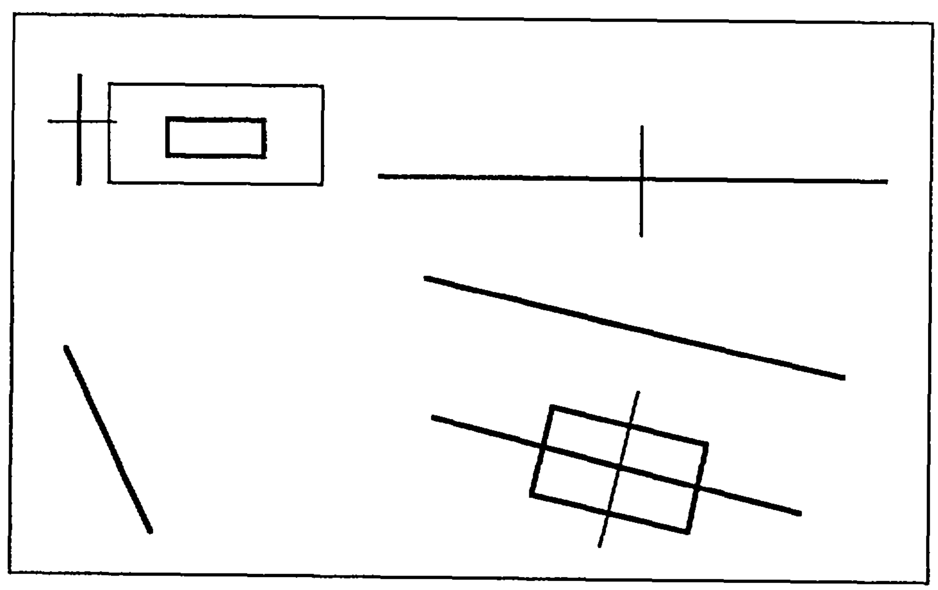

Fig. 4. Depictive search. Here various combinations of the shape depictors are used to depict properties of shapes in the watercolour landscape painting of Lyme Regis (Fig.5). In the simplest cases the depictors at the bottom left and middle right of the query represent only the length, orientation and position of the corresponding shapes in the picture. In the cluster at the the bottom right of the query all the attributes of a shape that can be represented using the system (eg length, width, orientation, size, and position, with tolerances) are combined.

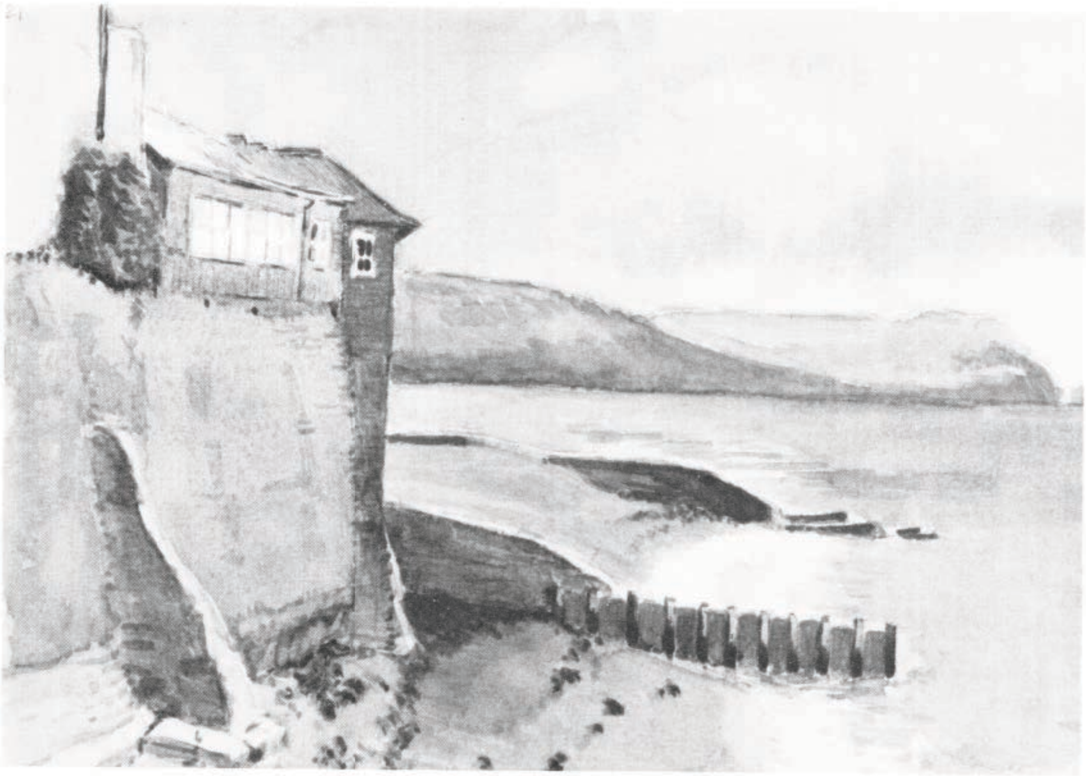

Fig 5. Stephen Scrivener, *Lyme Regis* , watercolour, 200 x 140mm, 1985. This picture is the basis for the depictive queries illustrated in Figs. 4 and 6.

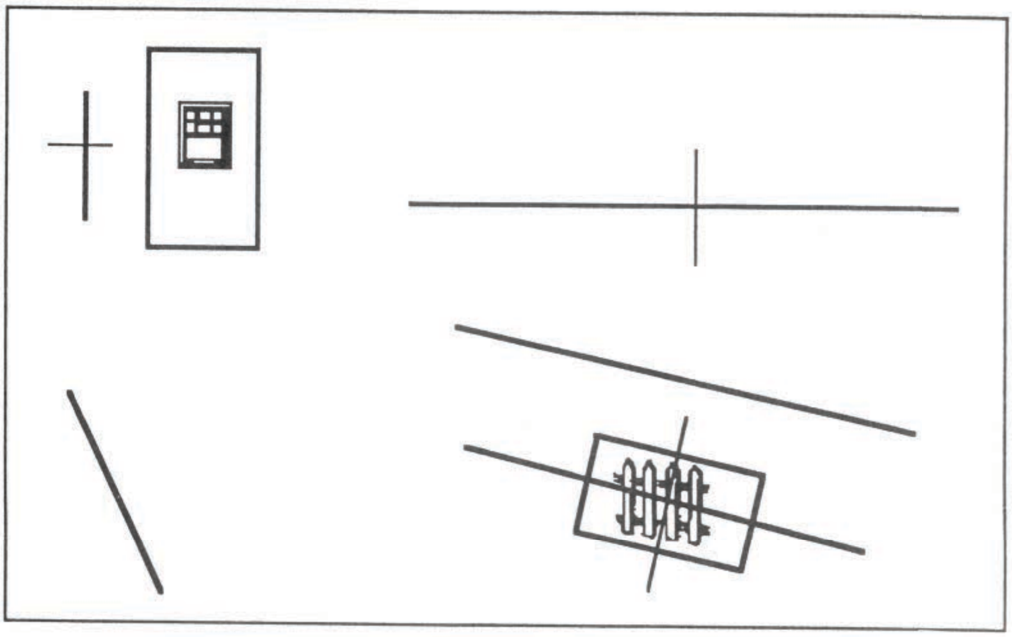

Fig 6. Decriptive-depictive search. Here the query representing the watercolour painting of Lyme Regis (Fig. 5) is augmented by the use of descriptive icons. It might be argued that these icons are depictive with respect to particular instances of the real world window and fence, or barrier, object types. However, with respect to the visual-spatial properties of *Lyme Regis* (Fig. 5) they are not depictive and are descriptive symbols for their respective object class. As such they are used as an alternative to writing.