Knowledge based interpretation of images: a biomedical perspective

NICHOLAS WALKER and JOHN FOX
Imperial Cancer Research Fund Laboratories, 44 Lincoln's Inn Fields, London, WC2A 3PX, UK

Abstract

The traditions of image processing and knowledge engineering have developed separately. Work on AI vision systems lies between the two traditions but only recently has attention been given to combining practical imaging systems with methods for exploiting knowledge in interpreting the contents of an image. Five general approaches to combining knowledge based expert systems with imaging technologies are discussed. Particular attention is paid to the requirement for techniques which transform a pixel array into a symbolic form suitable for interpretation, and current obstacles to a general solution. Interpretation of biomedical images is particularly problematic because of statistical, structural and temporal variation in morphology of objects and structures. Some ways in which knowledge of shape, structure, and object classifications may contribute to this interpretation are discussed. The survey focuses on biomedical images but many of the issues are of general relevance to work in image understanding.

1 Introduction

Images are an increasingly important source of diagnostic information in medicine. X-rays, and now computerised tomography, magnetic resonance, ultra-sound and nuclear isotope scanning techniques, provide two- and three-dimensional images, and even temporal sequences of images. Computers are routinely used to capture, generate and process these images; several computer science traditions have become involved in processing the images for medical decision support (e.g. see Viergever and Todd-Pokropek, 1988).

Most routine work in biomedical imaging, and indeed most other types of practical imaging, draws on mathematical signal processing techniques which make little if any explicit use of knowledge. This paper outlines some of the more obvious characteristics of current imaging technology; contrasts it with work in AI vision, and examines ways in which techniques from knowledge based systems might be used to augment conventional image analysis. After introducing the two major traditions in machine vision that have emerged in the last two decades, *traditional mathematical imaging* and *AI vision*, we illustrate some roles that knowledge-based systems could play in providing a new generation of flexible image understanding systems. We then discuss five general methods for coupling imaging devices to knowledge based systems, and finally discuss how knowledge of shape, structure and object taxonomies could be used to aid heuristic search of images, and the classification of objects found in them.

1.a Traditional imaging

1.a.1 Image processing

This tradition was one of the first to be established (a good introduction is in Gonzalez and Wintz, 1977). Image processing has been primarily concerned with developing mathematical methods for improving image quality in order to facilitate assessment by eye. Some imaging devices produce poor quality results because of intrinsically low signal to noise ratios or low spatial resolution. In medicine for example, gamma cameras—scintillation counters which detect emissions from radioactively labelled pharmaceuticals are widely used. Typically a gamma camera produces an array of about 128 × 128 pixels (picture elements), which can be treated as an image representing the spatial distribution of emissions due to organs or tissues taking up the pharmaceutical material at different rates. Techniques for selectively filtering unwanted noise, such as low pass filtering in the spatial frequency domain, or selectively enhancing statistically abnormal features are well established for use in gamma camera and other imaging systems. Enhancement of high spatial frequencies, band pass filtering, deblurring and many other techniques produce images which may be easier for a human observer to interpret.

1.a.2 Pattern recognition

These techniques extend the role of the computer into the realm of interpretation. Quantitative assessments or comparisons of image parameters can provide useful information which may not be immediately apparent to a clinician looking at a raw or even enhanced image. Additional statistical techniques can be used to exploit these quantitative parameters in order to classify images into clinically important categories (a simple example would be whether they are normal or abnormal).

1.b Image understanding

A common feature of medical imaging work, and most fields of imaging in fact, is a focus solely on the image—on improving its quality, extracting useful quantitative data from it and so on. Relatively little attention has been given to bringing non-image information to bear on describing the structures in the image or the meaning of those structures. In medicine, for example, the patient's medical history and a general knowledge of anatomy may be needed to determine whether an image feature is an unusual but normal structure, or whether it is a lesion.

Image understanding differs from imaging processing principally in the attempt to *describe* the world which is viewed through the imaging device. By this we refer to a number of capabilities, which include:

- the *articulation* of structures, objects or events within the image
- *recognition* of those entities as members of generic classes of structure etc.
- creation of a *symbolic data structure* which can support reasoning about the organisation and behaviour of objects; how they are causally or functionally related, and so on.

A number of restricted image understanding systems have been developed in AI. ACRONYM, for example, was specifically designed to recognise images of aircraft, though it was constructed as a general purpose vision system using "generalised cylinders" as a symbolic representation of the objects in the image and their parts (Brooks, Greiner and Binford, 1979; Brooks, 1981). VISIONS was designed to interpret outdoor colour scenes by identifying distinct regions (Hanson and Riseman, 1978) and SIGMA was designed to look for roads and houses in aerial photographs. (Matsuyama, 1984).

Most image understanding work has been linked to robotic manufacturing, satellite image interpretation or military recognition tasks. While the aims are exciting, current systems are quite limited. The restricted range of domains, and the relative poverty of the symbolic representations, means that these systems are collectively capable of addressing few tasks and individually they are not very versatile. Certainly nothing yet built would qualify as a *general purpose vision system*. When good performance is achieved on a task it generally depends on strong explicit or implicit constraints from knowledge of the domain; this works only so long as the application domain and task repertoire are limited.

2 Knowledge based expert systems for imaging?

To date, expert systems have been more concerned with decision making than with signal or image processing but they probably offer important ways of combining disparate kinds of information and increasing the versatility of image understanding systems. Knowledge based systems typically use knowledge of tasks, application domains, users and their requirements, their own structure and performance etc. to achieve explicitly specified goals.

Although a great deal of traditional image processing work has been carried out in biomedical fields, very little work in the AI tradition has been carried out. Expert systems for such tasks as diagnosis, treatment planning and so forth have of course received great attention in the medical world in the last few years and the medical imaging community has begun to ask whether, and in what ways, these techniques might be helpful to them. We shall first summarise the ideas introduced in expert systems which are most obviously relevant to image understanding problems, and then turn to schemes whereby they might be incorporated in flexible image interpretation systems.

2.a Qualitative representation of image structure

Perhaps the part of AI which has had the strongest influence on the emergence and development of expert systems has traditionally dealt with such topics as inference, problem solving, and the formal representation of intuitive human knowledge. These varied interests have converged in the development of techniques for qualitative and logical reasoning.

To extend techniques and tools for machine reasoning to reasoning about images we shall need to be able to represent qualitative aspects of the relationships among objects in a visual scene. We may think of this as like reasoning with a *map* of features and objects rather than an *image*. The precise geometry of the objects and their relationships may not be recorded; geometry is not a good representation for reasoning and, as illustrated by maps like maps of underground railways and stations, it may be relatively unimportant when compared with knowing what order the stations are in on a line, which stations are directly connected and which are not, and how you get from station A to station B. Similarly we may be more concerned with the temporal relationships between events (what happens before, during or after what) than the accurate chronometry of a sequence of photographic frames.

Numerical data are used by knowledge based expert systems (e.g. degrees of certainty, strengths of causality in diagnostic systems) but they are invariably used within a qualitative framework. Frequently however it may not be possible or advantageous to determine or estimate numerical parameters for qualitative descriptions, so wholly qualitative methods still require development. In *qualitative process theory* (Hayes, 1983; Forbus et al., 1984) qualitative relationships are expressed between quantities (e.g. the pressure in a fluid *increases with* depth) instead of modelling physical processes with sets of specific differential equations.

We therefore need to encode images, at least in part, in qualitative terms. As we have seen we are likely to need to represent spatial structures in two or three dimensions using topological descriptions rather than geometric coordinates, and temporal descriptions rather than chronometric ones. Examples of qualitative relationships which may be useful for topological reasoning about visual structures are *connectedness*, *adjacency*, *handedness* and *direction* (see Herskovits, (1986) for a discussion of spatial relations, and Rawlings et al. (1985) for an example of the use of a logic based spatial representation for reasoning about molecular structures) and terms like those introduced in temporal logics for reasoning about relationships between events (Allen, 1982) such as *before*, *during*, *simultaneous* etc.

2.b Explicit representation of knowledge

Conventional software encodes knowledge implicitly in algorithms or abstractly as numbers. Knowledge based systems make facts and programs explicit in symbol structures that can be searched and modified as well as executed. For image interpretation, facts like *region 1 is inside region 2* need to be encoded symbolically in such a way that the machine has access to information which defines what a "region" is, and what it means for one thing to be "inside" another and so forth.

We are also likely to need explicit techniques for representing abstract (non-spatio-temporal) relationships. Among the more obvious of these in medicine are representations of causal mechanisms underlying diseases; knowledge about the function of organs and physiological or biochemical systems; knowledge about the taxonomic relationships among diseases, symptoms, and so on. These will presumably be as important for interpreting images as interpreting information entered in other ways.

In a pure expert system (a pure system has probably never been implemented but no matter) *all* domain knowledge is explicit. Even procedural knowledge should be explicitly encoded in the sense that the system can examine imaging procedures, decide when to use them, and even modify or adapt them for particular circumstances. In many realistic image processing systems though, low level image processing algorithms will often have to be implemented as black boxes; algorithms which the system can apply to image data but cannot monitor or control in detail. However we can build a limited declarative representation of such processes: how long does each take to execute; how well does it work in different circumstances etc. We can then reason about the use of such algorithms using this explicit representation.

2.c *Strategic and meta-level knowledge*

The first wave of research into medical expert systems has been primarily concerned with learning how to represent knowledge in an explicit logical form so that it can be exploited in making decisions and solving problems. In most early programs the strategies for decision making and problem solving are incorporated into the code of the expert system software and are not made explicit. Recently, interest has turned to explicitly representing the decision making and problem solving strategies which underly good clinical judgements (see Davis and Buchanan, 1984). When should we try to eliminate possible diagnoses rather than confirm specific hypotheses? How much diagnostic detail should be collected before committing to treatment? How should the clinician cover for several possible but uncertain conditions? The ability of a machine to reason at the meta-level about its own knowledge and methods offers both an important source of flexibility and a way of formalizing knowledge based theories of decison making (Fox, 1984, 1986; Davis and Buchanan, 1984).

It may also be possible to explicitly represent tasks which are routinely required in interpreting a class of images, and to manipulate these representations at the meta-level. For example to carry out a diagnosis task based on an image, standard steps may be required including cleaning up the image, searching for gross abnormalities, considering whether any such abnormalities might explain clinical findings, describing the abnormality in detail, and so forth. The steps involved in such tasks, any constraints on their order of application, criteria for generating and evaluating diagnostic hypotheses etc. can all be represented explicitly in a knowledge base of a general medical decision support system (e.g. O'Neil, Glowinski and Fox, 1988). An attempt to explicitly represent some of the tasks which need to be executed to answer queries on biomedical images is described by Walker and Buxton (1988).

2.d *The user interface*

Finally a feature which is often considered vital in designing medical decision support systems, and which is in large part a consequence of making data, facts and meta-level knowledge explicit, is the ability for the user to engage in a question-answer dialogue with the system about what it knows, how it has arrived at its conclusions, whether such and such a hypothesis is supported, and so on (e.g. Fox, Glowinski and O'Neil, 1987). Questions from the user may sometimes be answered simply by retrieving information stored in the knowledge base—an expert system knowledge representation can be highly intelligible. In other cases a user's questions may cause the expert system to reason out the answer, or to set itself a task to collect additional data. These features encourage the view of an expert system as an active partner in clinical decision making; a significant change from the imaging community's traditional view of the imaging system as a mere instrument. The image interpretation system being developed by Walker and Buxton (op cit) sets out to provide a flexible user interface of this kind.

3 Knowledge based image interpretation—an illustrative example

Medical images are just one form of patient data. How could expert systems be used to help interpret them? Since expert systems offer the possibility of integrating many different types of knowledge and using that knowledge flexibly, they seem to offer some attractive possibilities. We now give an informal illustration of how general medical knowledge of the clinical presentation of different kinds of lesions and their typical anatomical locations could be used by a knowledge based system to generate hypotheses about what to expect in an image. The example illustrates how various sorts of knowledge might be used to direct the interpretation of a brain scan*.

Given the following proposition or fact:

symptoms of cerebellar tumour include dizziness

and the patient data:

history of J Smith includes dizziness

the following rule:

if history of Patient includes Symptom
and symptoms of Disease include Symptom
and Disease of Patient is not eliminated
then possible diagnoses of Patient include Disease

will generate the following conclusion:

possible diagnoses of J Smith include cerebellar tumour

Given a set of possible diagnoses generated from patient data we can now try to *verify* hypotheses on the image data. This is the classic generate-and-test strategy, and is probably the commonest approach to signal processing in AI. It can be conveniently illustrated using control-rules to specify a strategy for processing, encoding or searching images in ways which depend upon the hypothesis which is being pursued (or to reflect other goals). Given the above conclusion and the following facts:

- common locations of cerebellar tumour include posterior fossa
- morphologies of cerebellar tumour include dense sphere

Then this rule will be executed:

if possible diagnoses of Patient include Disease
and common sites of Disease include Site
and morphologies of Disease include Morphology
then verify presence of Morphology at Site

the following goal is generated:

verify presence of dense sphere at posterior fossa

* For intelligibility we express our example in pseudo-English, the representation language of Props 2, a general knowledge engineering package developed in the authors' laboratory. Capitalized strings are logical variables which are universally instantiated during inference.

In passing we might note that the examples are written in a *general* form; they are not specific to particular domains, particular medical conditions or imaging devices. Consequently as more medical facts are added to the knowledge base the scope of application of the rule set increases, but without making potentially troublesome alterations to the rule base and hence the logic of the inference strategy. Many expert systems depend upon knowledge bases which are built entirely out of special case rules; this limits the scope for identifying general strategic knowledge. In image interpretation as in other areas we expect a general trend towards representing knowledge in generalized logical form because of the power, flexibility, and scope for formal analysis which this offers.

Many knowledge sources could guide the generate and test process. For example image analysis could yield some features which are easily interpreted in light of clinical information, while others are not easily explained unless other findings could be confirmed on the image. The following rule is another strategic knowledge source:

If unexplained features include Feature
 and Feature is present at Site
 and Feature is possible part of Morphology
then verify presence of Morphology at Site

We now come full circle and generate a set of possible interpretations from what we have discovered about the image:

If Morphology is present at Site
 and morphologies of Disease include Morphology
 and current patient is Patient
 and Disease of Patient is not eliminated
then possible diagnoses of Patient include Disease
 and site of Disease is Site

An outstanding question concerns the recognition of objects and events; what does it mean, for example, to "verify" a hypothesis on an image? We shall return to this important question but for now note that we are, in effect, using our knowledge to build a model (or probably a set of models) of the objects of interest in the image. A *hypothesis* is a partially instantiated model which we can attempt to match in the image. Matching will either succeed (a structure in the image is a good match for the partially instantiated model and thereby completes the instantiation) or fail (assumptions of the model which constrain the structures or parameters of the image are violated) or no conclusion is possible. We shall consider how some types of knowledge can be used to control the recognition and verification processes after the next section, which discusses the prior question of how we might couple algorithmic operations on images with knowledge-intensive symbolic processes.

4 Coupling image processing to image interpretation

In this section we discuss five general ways in which an imaging system could be coupled to the inference mechanisms of an expert system. All the schemes assume two common components whose details we shall ignore since they are not germane to a general survey.

First, there will be some sort of signal processing device which delivers a quantized representation of the input signal as an array of pixels. We shall mostly consider just two-dimensional images here and occasionally images that change in time, but it should be remembered that real images are normally projections of three dimensional structures.

Second, we assume a knowledge based system which can make logical inferences about clinical significance on the basis of image features or structures once these have been extracted from the pixel array. The knowledge base might incorporate knowledge about parameter values; the appearance of particular types of lesion; the clinical symptoms and signs that are associated with

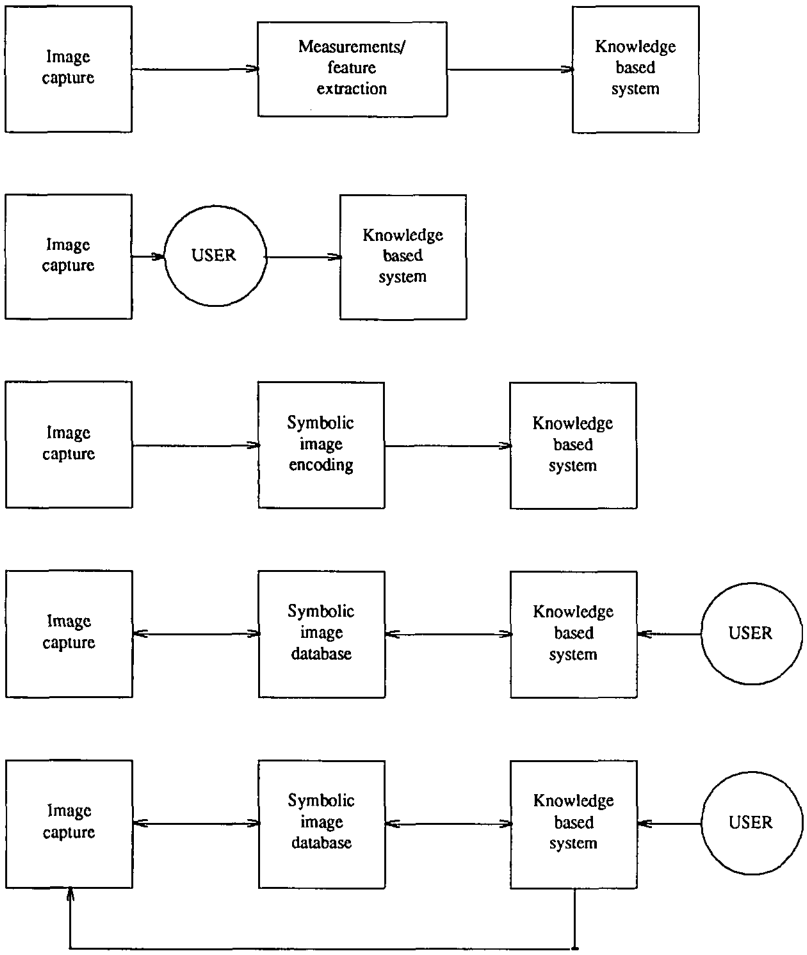

Figure 1 Five general schemes for coupling imaging devices and knowledge based expert systems. Details given in text

different lesions, and so forth. Again we shall not concern ourselves with specific applications, nor with technical commitments about how the knowledge should be represented, whether as rules, frames or otherwise.

What general techniques might one envisage then for coupling the pixel output of the signal processor to the symbolic input of the expert system? Complex, multi-pixel structures (such as contours, patches and textures) or organisations of structures (e.g. a dark patch adjacent to a curvilinear contour; an increase in patch diameter over time) are only implicit in a picture, and hence in a pixel array. They must be made explicit before the expert system can do any useful work. What general schemes are there for extracting features or structures from the pixel array and presenting them to the expert system?

Scheme 4.a represents an apparently simple coupling of the two subsystems using conventional pattern recognition procedures. Scheme 4.b requires a human participant to encode image features or structures while 4.c, 4.d and 4.e demand additional programs—currently beyond the state of the art—for creating and managing a symbolic description of image contents. Let us consider some of the characteristics of these schemes in more detail.

4.a Quantitative coupling

In this simple scheme conventional image processing algorithms are applied to the pixel array to yield (1) singleton parameters (e.g. average pixel value; DC level in the spatial frequency domain)

(2) distributions (e.g. a set of values of cross-correlations with images in a library; the power spectrum of the image) (3) derived boolean descriptors (e.g. normal/abnormal). Such data can now be passed for interpretation by an expert system prepared using one of the many expert system tools which are now available.

Given the output from an initial set of general algorithms, the expert system could then use its knowledge to select further procedures to further process and restrict the range of possible image interpretations. We return to the question of feedback from interpretation to image processing in a later section.

This may be a practical scheme for images when the clinical significance of particular parameters is known. However, it is obviously a limited application of AI technology. Indeed the approach may not warrant the use of a knowledge based system at all since the image representation is so impoverished. Since it does not describe the image in natural terms—objects and their parts; organs, tissues, lesions and their spatial topology etc.—the opportunity to use the spatial, temporal or causal forms of reasoning that knowledge based systems offer is very restricted. Such a system certainly could not be said to "understand" the image; it could not offer medically intelligible explanations of its conclusions, for example.

This is not to dismiss quantitative coupling. It may well be a practically valuable extension to existing pattern processing techniques, particularly as some of the more exotic possibilities of image understanding are still in the realm of AI research rather than practical development.

4.b Human encoding of image data

While people cannot perceive the power spectrum of an image, or reliably report its statistical properties, they can identify shapes and objects, areas of interest, possible abnormal formations, and so on. An expert system might therefore be developed to operate interactively with a human observer viewing the output of the imaging system. The expert system might be able to ask questions like "do you see any areas of high activity in the posterior portion of the scan?" and if the answer is positive, "is the area circular or elongated?", "is it close to the cranial wall?" and so on. Ellam and Maisey (1986) give an example of this approach applied to the interpretation of thyroid scans.

Some advantages and possible disadvantages of the approach are obvious. Obviously the scheme is straightforward to implement. Since an experienced observer will be able to make some judgements more effectively than a computer for some time, it is attractive to have an expert system which can interpret complex image features even it has no real understanding of them. On the other hand participation by skilled people demands expert human resources, which the system was at least in part intended to reduce. Furthermore, even skilled observers have lapses and make errors and human observers cannot easily quantify their observations; in this scheme the computer does not have access to the raw image data at all and so the system is losing quantitative and other information. As with scheme 1 this hybrid approach should certainly not be dismissed, but it is likely to be a transitional technique.

4.c Automatic symbolic encoding of the image

Scheme 1 is practical but depends upon pixel- and array-oriented computations on the image rather than on object-oriented computations which deliver an encoding in terms of organs and tissues and the relationships between them. Scheme 2 is also practical but has the opposite problem—an observer can describe objects and their spatial relationships but the underlying pixel data are not available for quantitative analysis. Clearly we would prefer to have programs which can accept pixel arrays and output explicit symbolic descriptions.

AI work in image understanding has developed some way towards providing general techniques for describing images. A great deal of effort has been expended in developing mathematical operators for finding edges and regions which perform well and are robust in the presence of noise, for example. Description of the important edge finders are available in standard texts on the subject and so are not covered here. All such operators however, whether region or edge based, only produce a data structure that is still strongly akin to an image rather than an interpretation. It is a heavily compressed encoding and more closely related to real aspects of the world (reflecting surface discontinuities; the boundaries of organs etc.) than to the original pixel array, but no interpretation in terms of the objects or events that produced it can be obtained directly.

Clearly we need something more, a symbolic description. A symbolic description is an interpreted, explicit representation of the image, with all the observed parts explained and perhaps the implicit aspects of it made explicit (such as the unobserved far side of objects). So to understand an image we have first to understand the symbolic description.

A way out of this impasse may be to build up symbolic descriptions from the primitive components (edges, regions etc.), first identifying "islands" or structures in the image which the program can be relatively confident have been interpreted correctly, and use these to constrain the interpretation of more ambiguous structures in the image. We may not need to generate a complete description fortunately, since the image may provide features which strongly constrain the number of interpretations which could explain them.

With simple features (e.g. corners, circular arcs, or combinations of these and edges) we can be fairly sure of their identity though a large number of hypothetical objects could give rise to them in the image. If a feature is complex it is harder to describe, but a more limited number of explanations are possible.

In a non-medical case ACRONYM (Brooks, 1981) looks for instances of a "ribbon" feature—two matching edge segments lying on either side of a hypothetical axis—which is taken to be the two-dimensional projection of a cylinder. More complex features can be generated by processes which find relationships between primitives that are unlikely to have arisen by chance. These processes are termed *perceptual grouping*, a term deriving from work on human perception. *Continuities* and *regular structures* in images (for example, parallel or collinear structures or those having an axis of symmetry) are enormously informative because they are so unlikely to have arisen by chance. Consequently they can be used to simultaneously dissect the image into related parts and those which are unrelated and should be attended to and interpreted separately (e.g. Fox, 1978). Continuity relationships also tend to remain highly invariant over large ranges in the angle at which an object is viewed and often over different instances of objects within a class. Lowe (1987) provides an example of this sort of system in SCERPO.

Automatic symbolic description of images is generally difficult and particularly so in medicine. Among the reasons for the special difficulties in medicine are:

- Current feature extraction techniques are not very robust while medical images are frequently of low resolution and very noisy. This situation is currently improving with improvements in imaging devices, but it seems clear that it is not possible to extract features effectively without a heavy computational overhead. Happily this can be effected in a highly parallel way and specialized chips are becoming available which can carry out parallel extraction of simple features, like edges, across an image in near real-time.
- There is high statistical variability in image parameters and object geometry (lesion morphology and "normal" anatomical features). Biological tissues are frequently soft, so the shape they project on an image is highly variable, as both normal and abnormal processes affect their three-dimensional morphology. This produces great difficulties in producing a "natural" parameterisation of objects which allows recognition. Several approaches to this have been suggested. Pentland (1987), for instance uses a combination of fractals and superquadratics to provide a representation of such shapes.

We are some way from designing a system that can automatically produce a good symbolic representation of the shape of a non-rigid, poorly defined structure in the presence of other structures. This is of course the kind of image which is routine in, say, medical radiology.

4.d The image as a database

An assumption we might easily make about a symbolic image description is that it should explicitly contain all possible descriptions of the image, just as the pixel array contains all the information in the original two-dimensional signal (assuming the spatial sampling frequency is high enough). Although in some simple image understanding problems it may be possible to precompute all possible descriptions, this is not generally true.

Firstly it is well known that images are often underspecified, in the sense that there will be more than one consistent interpretation of the image (however high the fidelity of the pixel array); the standard illustration of this is that an infinite number of three-dimensional objects can project the same shape onto a two-dimensional image plane. Furthermore some ambiguities in the image can only be resolved by means of external information such as clinical knowledge about the kinds of lesion that are possible.

Secondly we cannot expect to have a complete symbolic description of the image because descriptions are not absolute but relative to a purpose or question—what a radiologist wants to extract from an image may well depend upon the diagnosis that is suspected, or the treatment that is being evaluated.

The image description should not therefore be viewed as a closed structure but as a database of descriptions which the expert system can search, and add to as new descriptions are derived, which are specific to the clinical questions at hand. Buxton and Walker (1987) describe experimental systems aimed at treating time-lapse cinemicrography film as a symbolic database of objects and events which can be queried in a variety of ways. In effect the system constructs a proof of the query, by performing only the processing construction and search of the image database which is necessary for answering that query.

For many queries the system can ignore large parts of the image as irrelevant to the task at hand, and only generate a minimal symbolic representation of them. A side effect of processing may be to update the symbolic representation. In order to know what is relevant and what is not, what representations to build, and when to stop improving a description or specializing a hypothesis, the program must have access to knowledge of what it means to satisfy a particular query. As discussed above this knowledge can often be encoded declaratively in another, non-image part of its database.

4.e Active control of imaging

The fourth scheme emphasized the ability to update the symbolic image database with information or inferred interpretations which reflect the context of a user's queries. The next obvious question is whether future expert systems will need to control image capture and low level processing. For example should it be possible to adjust the parameters of the imaging device, the quantization of the image, or the symbolic encoding of the image in the light of system goals or prior image analysis?

The requirement seems plausible for two reasons. First, imaging research has produced a wide range of operators for extracting edges, regions, textures, and so on. So far however there is no universal operator or operators; each has particular strengths and weaknesses. Unless one computes all possible operations which may be impractical, there is a risk of losing potentially significant information because the wrong operator is used. For example, different features have different sizes, and operators with different resolutions are needed to extract them (we shall discuss this again in the section on scale hierarchies below).

Human and many other biological vision systems are clearly designed to actively modify the image capture process. The human retina provides high resolution of detail in the central region, the fovea, with resolution decreasing dramatically towards the periphery. In contrast the periphery has high temporal sensitivity, for movement and flicker. We move our eyes strategically to capture image data, using the fovea to direct attention to interesting areas of a visual scene, track selected moving objects and so on, perhaps after detecting gross image features or changes peripherally. A radiologist scans a radiograph under the guidance of clinical knowledge (knowing what he/she is looking for) and the low resolution information coming in from the periphery. Similar benefits may accrue from active control of machine imaging.

The second argument is related but is concerned with image enhancement. When specific clinical questions are raised it may be useful to collect image data in a way that reflects these questions. For example high frequency information may mask structures recorded in the low spatial frequencies. If we are looking for structures which are expected in the low frequencies such as a large, fuzzy patch, then reprocessing the image with high spatial frequencies removed might be helpful.

This question of whether image capture should operate in a fixed way, and whether it should reflect the clinical task which is being carried out, seems to be a variant of the earlier argument that the database of symbolic descriptions is not closed. The set of measurements we make on the image should not necessarily be viewed as closed either, because the choice of measurements makes implicit assumptions about what we are making the measurements for. If a "neutral" set of operators cannot be found, and the image capture process has to reflect idiosyncrasies of the clinical situation, then the expert system may require two very different kinds of knowledge (1) *domain* knowledge about the clinical significance of particular structures and (2) *imaging* knowledge which embodies know-how about imaging itself (e.g. see Nazif and Levine, 1984 for a system where the very low level operations are made explicit as production rules and thus represent a body of knowledge about imaging).

As a concrete example of the potential value of active control of imaging consider magnetic resonance imaging. The images obtained by this technique depend on the particular parameters of a magnetic pulse sequence. A body of knowledge is being accumulated about the pulse sequences which produce the best diagnostic accuracy for different pathological processes. A knowledge based system which has access to (1) clinical information on the patient; (2) medical diagnostic knowledge; (3) the body of knowledge about magnetic resonance techniques and (4) the results of a trial image perhaps, could be combined to determine an appropriate pulse sequence to answer specific clinical questions.

5 Symbolic encoding and representational hierarchies

Having looked in general terms at ways in which we might couple existing imaging devices with knowledge based system software we now return to the central question of how knowledge could actually be *used* in the image interpretation process.

A fundamental principle which has emerged recently from work in image understanding is the value of using a range of different types of knowledge at all points in the signal-to-symbol transformation. Heuristic knowledge, which is frequently represented in categorical hierarchies, can be used to constrain the interpretations which are assigned to structures which have variable geometry. Vision systems which use such non-image heuristics are likely to be more robust than those which don't. The following examples are not comprehensive but illustrate the ideas involved.

5.a Scale hierarchies—scale in the image

Images contain information at a broad range of spatial resolutions or *scales.* The lungs in a chest X-ray, for example, can be described as containing specific detailed structures when looked at closely, as a textured field when this detail is disregarded, or as a grey area of a particular shape against a black background. An edge can be wavy at one scale but straight at higher scale (lower resolution). To recover the different types of information which are available at different scales image processing operators must have a range of sizes (or equivalently a range of pixel neighbourhoods over which their output is calculated). Operators with a large neighbourhood miss fine detail but are relatively insensitive to noise, while the converse is true for small neighbourhoods. The result of using a range of operator sizes is an explicit representation of features at different scales—a *multiresolution* representation. An image and its representation at four different scales is illustrated in figure 2.

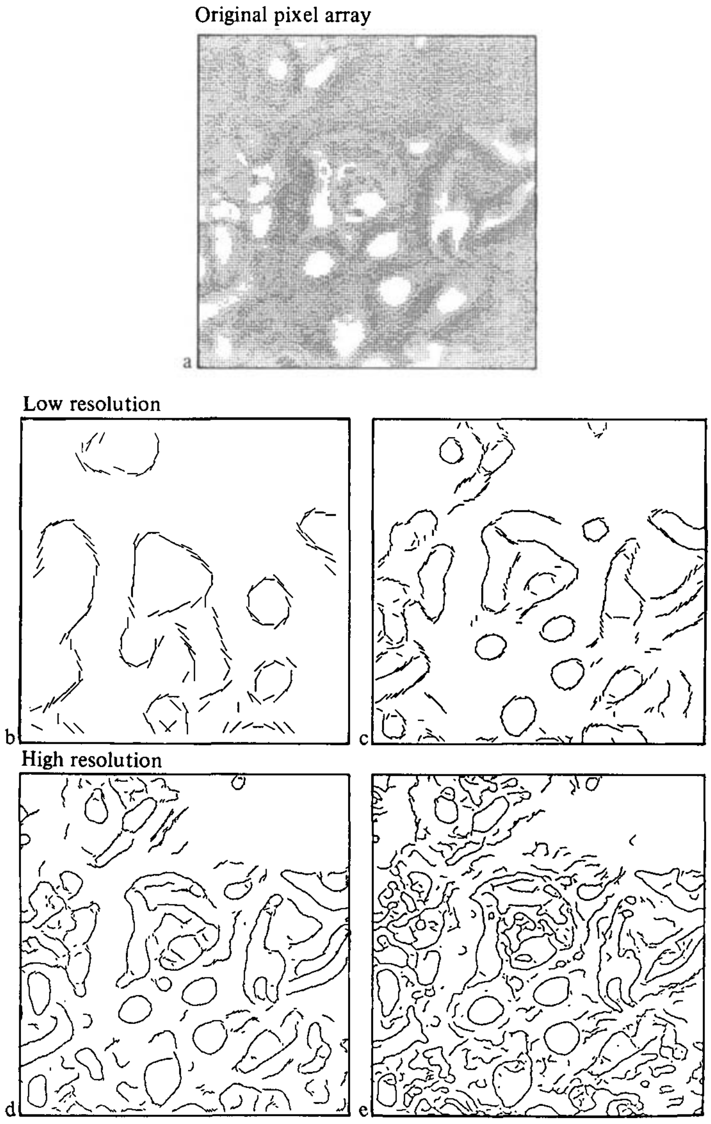

Figure 2 A multiresolution view of a complex aggregate of cells.

Current vision systems (e.g. experimental robot vision systems) typically operate at a single scale. By doing so they implicitly make strong, simplifying assumptions about the nature of the world; in effect they assume that it is populated by smooth, rigid objects at a known distance from the point of observation. This is computationally convenient and realistic for many industrial applications. Biomedical applications however, require the capability to deal with rough, non-rigid things at different depths. At any spatial resolution which is coarser than the "natural scale" of the image, irregularities of surfaces are smoothed out. Analogously, if dynamic images are examined over a short enough time interval (high temporal resolution) even highly plastic objects will be locally rigid. Since much of the significant information lies in spatial and temporal structures whose scales are variable any single resolution will frequently prove to be inappropriate.

A multiresolution approach, which provides a capability to search through a range (hierarchy) of scales in order to verify or eliminate the presence of expected structures, may allow vision systems to exploit simplifying assumptions without being overcommitted to them. This is likely to be particularly important in biomedical image interpretation. In fact, low level processing algorithms are often inherently multiresolutional, though little exploitation of the potential of multiresolution processing has been achieved to date. Work on biomedical images, we believe, may give an important boost to the development of flexible vision systems which can operate over a range of scales precisely because they do not make it easier for designers to escape hard problems in vision theory, by adopting simplifying, application-specific assumptions.

The next two sections illustrate how the arguments for a multiresolution approach may also apply to computations which are carried out on hierarchies at higher symbolic levels, as in the characterization and recognition of the shape and part-structure of objects.

5.b Shape hierarchies—scale in the objects

Recognition is the successful match of an object description with a model description. In a multiresolution representation, objects and models will possess multiple descriptions at different scales. Recognition can then be treated as a successful match *at some resolution.* If we represent three-D objects with cylinder-like primitives for example (Marr and Nishihara, 1978; Brooks, Greiner and Binford, 1979), then a human figure consists of an upright elongaged blob at one scale, but at a higher scale it will consist of a round blob (head) on an elongated blob (trunk), with elongated blobs representing arms and legs. As we pursue more and more detailed descriptions we match object and model more closely, progressively achieving a more precise classification. At some point in our hierarchy of shape descriptions we recognize a human being, at another a female, and at another a particular individual.

The hierarchical representation of shape contributes to the robustness of the recognition process. For example, it is not very useful if our recognition system simply fails when it tries to match an object instance with a stored description, as might happen with a single resolution representation. It would be rather more flexible if it returned some information about where the failure occurs in the shape hierarchy, e.g. the object is a person, though the sex is undetermined, or the object appears like a round, dense tumour though its detailed morphology cannot be seen. With increasing noise, obscuring structures and so on, the shape hierarchy can be used to limit the possible mappings between the external object representation and the internal descriptions and hence constrain the interpretations. Once this has been done and we have extracted the maximal information we can from the image, we can then use other knowledge to provide further interpretation constraints.

An important problem discussed above in section 3, was what it might mean to *verify* the presence of an object or structure in an image. We can now make a start on a tentative interpretation: take a low resolution image, use the gross features extracted from it together with knowledge of what structures and objects might be expected in this type of image and generate a set of provisional hypotheses. The initial hypotheses are described at the "top" of the scale/shape hierarchy. Now we can "descend" the shape hierarchy, specify the shape description to be expected at a higher resolution, and search for matching features in the image database. The expected structure may be successfully verified at this scale. If not, then new, unexplained features may be found, which can generate new hypotheses to be verified.

5.c Part hierarchies

Parts of objects are objects in their own right, and they in turn have parts. Knowledge of the *part hierarchy* of objects is also likely to make important contributions to verification and recognition processes. Sometimes we can only recognize an object when we have recognized some of its parts. In many natural visual scenes objects are frequently partially occluded by other objects. In images of tissues, organs and so forth, parts of structures may be distorted by normal or disease processes.

Parts of objects or structures may be viewed as complex features. If a feature is recognized, then knowledge of the objects of which it may be a part, permit the vision system to hypothesize the presence of the parent object. The system can then search the image database for information to confirm the hypothesis. Even fragmentary recognition of an object constrains the possible labellings assigned to its parts, and vice versa. For example, among the criteria for identifying abnormal cells in cervical smears are abnormal numbers of cell nucleii and abnormal shape of the nucleus. The relative positions of objects and their parts, and the parts and the envelope of the parent structure, offer further constraints. These can be used to eliminate class descriptions from the search tree and constrain the interpretation of remaining elements. With increasing resolution not only does identification improve—the constraints on the position and orientation descriptions do also (Grimson, 1987).

These processes suggest an attractive way of combining bottom-up and top-down processing in recognition. Verification techniques which expoit scale, shape and part hierarchies can be viewed as variants of the process of *heuristic classification* described by Clancey (1986); a method which was used by many first generation expert systems to exploit taxonomic knowledge in refining hypotheses into progressively more specific forms.

6 Summary and conclusions

Over the last two decades or so great progress has been made in developing practical image capture and image enhancement techniques. Work in AI is relatively experimental but some techniques, particularly feature detection and grouping, knowledge and inference methods, have aroused the interest of the biomedical engineering community. These systems raise the possibility of combining image processing technology with clinical knowledge engineering for interpreting image data. Some aspects of expert systems technology which are most relevant to image interpretation have been outlined, notably the use of explicit and qualitative representations of knowledge, meta-level reasoning and advanced user interface methods.

Ways in which expert system concepts might be exploited in image interpretation have been illustrated, and five general schemes for coupling knowledge based systems to image processors have been outlined. Only two of these are currently practical; the remainder are restricted by our inability to create symbolic representations of the structures typically found in biomedical images. Some ways in which knowledge of shape, structure, and object taxonomy may contribute to the interpretation of images have been discussed.

Important areas for further research include (1) exploring the limits of existing symbolic encoding techniques; (2) identifying the optimal set of information types which should be encoded in a symbolic image description database; (3) developing symbolic representations of the characteristic spatial and temporal topologies of biological objects and events; (4) identifying ways in which non-image domain knowledge can be used to constrain the interpretation of image elements and (5) establishing the degree to which the low level image capture mechanisms needs to be controlled by knowledge-intensive interpretation processes.

Biomedical images raise special challenges due to the variability and non-rigidity of biological structures. Ironically the very difficulty of interpreting these images may contribute to the development of general purpose vision systems because the simplifying assumptions which can make specialized vision applications relatively tractable cannot be adopted in the biomedical domain.

References

Allen, J, 1984. "Towards a general theory of action and time" *Artificial Intelligence* **23**(2) pp. 123–154.

Brooks, R, Greiner, R and Binford, T, 1979. "The ACRONYM model-based vision system" *Proc IJCAI* **6** Tokyo 1979 pp. 105–113.

Brooks, R, 1981. "Symbolic reasoning among 3-D and 2-D images" *Artificial Intelligence I* **17** pp. 285–348.

Buxton, H and Walker, N, 1987. "Query based visual analysis: spatio-temporal reasoning in computer vision' Proceedings of Alvey Vision Conference, London.

Clancey, W J, 1986. "Heuristic classifcation" *Artificial Intelligence* **27** pp. 215–251.

Davis, R and Buchanan, B G, 1984. "Meta-level knowledge" In: *Rule-based expert systems*, Buchanan, B G and Shortliffe, E H, Reading, Massachusetts: Addison-Wesley Chapter 28.

Ellam, S V and Maisey, M N, 1986. "A knowledge based system to assist in medical image interpretation: design and evaluation methodology" In: *Research and Development in Expert Systems III*, Bramer, M, (Ed.), Cambridge: Cambridge University Press.

Fox, J, 1978. "Continuity, concealment and visual attention" In: *Strategies of Information Processing*, Underwood, G, (ed.), London: Academic Press.

Fox, J, 1984. "Formal and knowledge-based methods in decision technology" *Acta Psychologica*, **56** pp. 303–331.

Fox, J, 1986. "Three arguments for extending the framework of probability" In: *Uncertainty in artificial intelligence*, Kanal, L N and Lemmer, J F, (Eds.), Amsterdam: North Holland.

Fox, J, Glowinski, A J and O'Neil, M, 1987. "The Oxford system of medicine: a prototype information system for primary care" In: *Proceedings of European Conference on Artificial Intelligence in Medicine*, Fox, J, Fieschi, M and Engelbrecht, R, (Eds.), Berlin: Springer-Verlag.

Forbus, K D, 1984. "Qualitative Process Theory" AI **24** pp. 85–169.

Gonzalez, R C and Wintz, P, 1977. *Digital Image Processing*, London: Addison-Wesley.

Grimson, W E, 1987. "Recognition of object families using parametized models" Proceedings of First International Conference on Computer Vision, IEEE Computer Society Press, pp. 93–101.

Hanson, A R and Riseman, E M, (Eds.), 1978. *Computer Vision Systems*, New York: Academic Press.

Hayes, P J, 1983. "The second naive physics manifesto" *Representations of common sense* pp. 468–485.

Lowe, D G, 1987. "Three-dimensional object recognition from single two-dimensional images" *Artificial Intelligence*, **31**(3) 255–396.

Marr, D and Nishihara, H K, 1978. "Representation and recognition of the spatial organisation of three dimensional shapes" Proc Royal Soc London, B **200** pp. 269–294.

Matsuyama, T, 1984. "Knowledge organisation and control structure in image understanding" IEEE Proc International Conference Pattern Recognition (1984) pp. 1118–1127.

Nazif, A and Levine, M, 1984. "Low level image segmentation: an expert system" IEEE PAMI-6 No. 5 pp. 555–577.

O'Neil, M, Glowinski, A J and Fox, J, 1988. "Decision making in the Oxford system of medicine" Proc. Conference on Expert Systems and their Applications, Avignon, 1988.

Pentland, A P, 1986. "Parts: structured descriptions of shape" Proceedings of AAAI, Pittsburgh pp. 695–701.

Rawlings, C J, Taylor, W, Fox, J, Nyakairu, J, Sternberg, M J E, 1985. "Reasoning about protein topology using the logic programming language Prolog" *Journal of Molecular Graphics* 3(4) pp. 151–157.

Viergever, M A and Todd-Pokropek, A, (Eds.), 1988. *Mathematics and Computer Science in medical imaging*, Berlin: Springer.

Further reading

General reviews of image processing and interpretation

Binford, T 1982. "Survey of model-based image analysis systems" *International Journal of Robotics Research*, **1** pp. 18–64.

Brady, M, 1982. "Computational approaches to image understanding" *Computing Surveys* **14** no 1 pp. 3–71.

Chin, R T and Dyer, C R, 1986. "Model-based recognition in robot vision" *Computing Surveys* **18** no 1 pp. 67–108.

Gonzalez, R C and Wintz, P, 1977. *Digital Image Processing*, London: Addison-Wesley.

Hanson, A R and Riseman, E M, (Eds.) 1978. *Computer Vision Systems*, New York: Academic Press.

Marr, D, 1976. *Vision*, New York: Freeman.

Viergever, M A and Todd-Pokropek, A, 1988. *Mathematics and Computer Science in medical imaging*, Berlin: Springer.

Examples of experimental image understanding systems

Brooks, R, Greiner, R and Binford, T, 1979. "The ACRONYM model-based vision system" *Proc IJCAI* **6** Tokyo 1979 pp. 105–113.

Brooks, R, 1981. "Symbolic reasoning among 3-D and 2-D images" *Artificial Intelligence I* **17** pp. 285–348.

Lowe, D G, 1987. "Three-dimensional object recognition from single two-dimensional images" *Artificial Intelligence* **31**(3) pp. 255–396.

Ohta, Y, 1980. "A region-oriented image-analysis system by computer" PhD thesis Kyoto University, Dept of Information Science.

Matsuyama, T, 1984. "Knowledge organisation and control structure in image understanding" IEEE Proc Int Conf Pattern Recognition (1984) pp. 1118–1127.

Matsuyama, T and Hwang, V, 1985. "SIGMA, a framework for image understanding—integration of bottom-up and top-down analysis" Proc IJCAI 1985 pp. 908–915.

McKeown, D, 1985. "Rule-based interpretation of aerial imagery" IEEE PAMI-7 No 5 pp. 570–585.

Niemann, H, 1985. "A knowledge based system for analysis of gated blood pool studies" IEEE PAMI-7 No. 3 pp. 246–259.
Levine, M, Noble, P and Youssef, Y, 1983. "A rule-based system for characterising blood cell motion" In: *Image Sequence Processing and Dynamic Scene Analysis* Huang, T S (Ed.) NATO ASI series **2** Berlin: Springer Verlag pp. 663–709.

Low-level Vision

Canny, J, 1983. "A variational approach to edge detection" Proceedings of National Conference on Artificial Intelligence AAAI-83.
Marr, D and Mildreth E C, 1980. "Theory of edge detection" Proc. Roy Soc B 207 pp. 187–217.
Nazif, A and Levine, M, 1984. "Low level image segmentation: an expert system" IEEE PAMI-6 No. 5 pp. 555–577.

High-level vision

Connell, J H and Brady, M, 1987. "Generating and generalising models of visual objects" *Artificial Intelligence* **31**(2) pp. 159–184.
Grimson, W E, 1986. "Recognition of object families using parametized models." Proceedings of First International Conference on Computer Vision, IEEE Computer Society Press, 1987 pp. 93–101.
Marr, D and Nishihara, H K, 1978. "Representation and recognition of the spatial organisation of three dimensional shapes" Proc. Roy Soc London B 200 pp. 269–294.
Pentland, A P, 1986. "Parts: structured descriptions of shape" Proceedings of AAAI, Pittsburgh pp. 695–701.