COMPUTATIONAL IMAGERY

J.I.Glasgow and D. Papadias
Department of Computing and Information Science
Queen's University Kingston, Canada, K7L 3N6
janice@qucis.queensu.ca

1 Introduction

Numerous psychological studies have been carried out and several, often conflicting, models of mental imagery have been proposed. This paper does not present another computational model, but instead treats imagery as a problem solving paradigm in artificial intelligence. We describe a concept of *computational imagery* [Papadias & Glasgow, 1991], which has potential applications to problems whose solutions by humans involve the use of mental imagery. As a basis for computational imagery, we define a knowledge representation scheme that brings to the foreground the most important visual and spatial properties of an image. Although psychological theories are used as a guide to these properties, we do not adhere to a strict cognitive model; whenever possible we attempt to overcome the limitations of the human information processing system. Thus, our primary concerns are efficiency, expressive power and inferential adequacy.

Computational imagery involves tools and techniques for visual-spatial reasoning, where images are generated or recalled from long-term memory and then manipulated, transformed, scanned, associated with similar forms (constructing spatial analogies), pattern matched, increased or reduced in size, distorted, etc. A primary goal of our approach to computational imagery is to facilitate the retrieval of visual and spatial information that was not explicitly stored in long-term memory. The images generated to retrieve this information may correspond to representations of real physical scenes or they may be abstract concepts that are manipulated in ways similar to visual forms.

The knowledge representation scheme for computational imagery separates visual from spatial reasoning and defines independent representations for the two modes. Whereas visual thinking is concerned with *what* an image looks like, spatial reasoning depends more on *where* an object is located relative to other objects in a scene (complex image). Each of these representations is constructed, as needed, from a descriptive representation stored in long-term memory. Thus our scheme includes three representations, each appropriate for a different kind of processing:

- An image is stored in long-term memory as a structured, descriptive, *deep representation* that contains all the relevant information about the image.
- The *spatial representation* of an image denotes the image components symbolically and preserves their spatial relationships.
- The *visual representation* depicts the space occupied by an image as an occupancy array. It can be used to retrieve information such as shape and volume of image components.

While the deep representation is used as a permanent store for information, the spatial and visual representations act as working (short-term) memory stores for images.

A formal theory of arrays provides a meta-language for specifying the representations for computational imagery. Array theory is the mathematics of nested, rectangularly-arranged data objects [Jenkins & Glasgow 1989, More 1979]. Several primitive functions, which are used to retrieve, construct and transform representations of images, have been specified in the theory and mapped into the functional programming language Nial [Jenkins et al. 1986].

2 Mental Imagery

Although no one seems to deny the existence of the phenomenon called imagery, there has been a continuing debate about the structure and the function of imagery in human cognition. The imagery debate is concerned with whether images are represented as *descriptions* or *depictions.* It has been suggested that descriptive representations contain symbolic, interpreted information, whereas depictive representations contain geometric, uninterpreted information [Finke et al. 1989]. Pylyshyn, a forceful proponent of the descriptive view, argues that mental imagery simply consists of the use of general thought processes to simulate perceptual events, based on tacit knowledge of how these events happened [Pylyshyn 1981]. He disputes the idea that mental images are stored in a raw uninterpreted form resembling mental photographs and argues for an abstract format of representation called propositional code. Kosslyn's model of mental imagery [Kosslyn 1980] is based on a depictive theory which claims that images are quasi-pictorial; that is, they resemble pictures in several ways but lack some of their properties. According to Kosslyn's model, mental images are working memory, visual representations generated from long-term memory, deep representations. A set of procedures, which is often called the "mind's eye", serves as an interface between the visual representations and the underlying data structures, which may be decidedly non-pictorial in form.

Findings, provided by the study of patients with visual impairments, have demonstrated that there are two distinct cortical visual systems [Mishkin et al. 1983]. This research indicates that the temporal cortex is involved in recognizing *what* objects look like, while the parietal cortex determines *where* they are located. More recent studies have also suggested that there exist two distinct components of imagery, the spatial and the visual, where the spatial component preserves the *where* information and the visual preserves the *what* information [Kosslyn 1987].

In this paper, we do not propose a complete model for imagery. Rather, we present a knowledge representation scheme that attempts to capture the underlying principles of imagery. The representation scheme extends Kosslyn's computational model by considering images as potentially *3D* and hierarchical. Our approach also supports two independent working-memory representations for imagery, corresponding to the visual and spatial components of mental imagery.

3 Knowledge Representation

We define computational imagery as the ability to represent, retrieve and reason about information not explicitly stored in long-term memory. In particular, we are concerned with visual and spatial information. Recall that the visual component of imagery specifies how an image looks and is used to retrieve information such as shape, size and volume, while the spatial component of imagery denotes where components of an image are situated relative to one another and is used to retrieve information such as neighborhoods, adjacencies, symmetry and relative locations. The long-term memory representation is implemented as a description of the image and the working memory representations correspond to representations that make explicit the visual and spatial properties of an image. In the remainder of this section, we discuss each of the representations and the primitive functions that operate on them.

3.1 Deep Representation

The deep representation is used for the storage of images in long-term memory. Most of the research in vision and imagery has focused almost exclusively on the format of the on-line conscious representations, to the exclusion of that entailed in long-term storage. Our point of view is that the deep representation falls more in the limits of research in long-term memory than imagery and we base its implementation on the hierarchical network model of semantic memory [Collins, Quillian 1969]. This model is suitable for storing images since they have a structured organization in which subimages can occur as elements in more complex images.

The deep representation in our scheme is implemented using a frame language, in which each image frame contains all the salient information about the image. This information includes propositional and procedural knowledge, including encodings that permit the reconstruction of the working-memory representations. There are two kinds of image hierarchies in the scheme: the AKO (a kind of) and the PARTS. The AKO hierarchy provides property inheritance: images can inherit properties from more generic image frames. The PARTS hierarchy is used to denote the structural decomposition of complex images. The deep representation for imagery, as most frame representations, can be characterized as non-monotonic, since default information (stored in specific slots, or inherited from more generic frames) might be superseded as new information is added to a frame.

Our implementation of the deep representation has several attractive properties. First it provides a natural way to represent knowledge since all the information about an image (or a class of images) can be stored in a single frame and the structure of images is captured by the PARTS hierarchy. It incorporates the psychological concept of semantic networks in a computational implementation that provides features such as conceptual hierarchies and procedural attachment. The non-monotonic feature of the frame structure matches the cognitive ability to make conjectures and infer information when the existent knowledge is incomplete. Despite its attractive properties, the deep representation is not the most suitable representation for all of the information processing involved in imagery. Thus, we require alternative representations to facilitate the efficiency of the scheme.

3.2 Visual Representation

Mental images are not constantly experienced. When an image is needed, it is generated on the basis of perceived or stored information. Thus, unlike the deep representation, which is used for the permanent storage of image information, the working memory representations exist only during the time an image is active, i.e., when visual or spatial information processing is taking place.

The visual representation corresponds to the visual component of imagery, and it can either be reconstructed from the underlying deep representation or generated from low level perceptual processes. Similar to Kosslyn's skeletal image [Kosslyn 1980], this representation is depictive and incorporates geometric information. Unlike Kosslyn's approach, we assume that the visual representation can be *3D* and viewer-independent.

For the current implementation of the visual representation we use *occupancy arrays.* An occupancy array consists of cells, each mapping onto a local region of space and representing information such as volume, lightness, texture or surface orientation about this region. Objects are depicted in the arrays by patterns of filled cells isomorphic in surface area to the objects.

Storing occupancy arrays in memory can be a costly approach. As a result other approaches to storing or generating this information (like quadtrees or generalized shapes) have been developed. Any of these approaches could be incorporated into a particular application of the scheme for computational imagery.

3.3 Spatial Representation

A primary characteristic of a good formalism for knowledge representation is that it makes relevant properties explicit. While an occupancy array provides a representation for the visual component of imagery, it is basically uninterpreted. For the spatial component of imagery we are best served by a representation that explicitly denotes the spatial relations between meaningful parts of an image. Thus we use a multidimensional symbolic array to depict the spatial structure of an image, where the symbolic elements of the array denote its meaningful parts [Glasgow 1990]. The symbolic array captures the spatial and topological relationships of the image features, but not necessarily relative sizes or distances. The spatial locations of the elements in the arrays can be interpreted in different ways depending on the domain of interpretation. If, for example, we use the scheme to reason about mental geographic maps, interpretations could include predicates such as *north, east, south* and *west*; if the array is used to represent the image of a room, then the interpretation would involve predicates such as *above, behind, left-of, beside,* etc. The spatial representation can also denote non-spatial dimensions. For example, the symbolic array could be used to order and index features such as height or speed.

The symbolic array representation for the spatial component of imagery is generated, as needed, from information stored explicitly in the frame representation of an image, or as a result of a successful image analysis. Note that parts of a spatial image may occupy more than one location in an array. This may be necessary to capture all the spatial relationships. For example, we may wish to represent that one component is located "inside" another, such as water in a glass. This can be accomplished with a symbolic array in which the symbols for glass enclose the symbol for water:

glass	water	glass
glass	glass	glass

According to Pylyshyn, images are not raw uninterpreted mental pictures, but are organized into meaningful parts, which are remembered in terms of their spatial relations. When we forget a part in an image it is not a random, but a meaningful part. Furthermore, we can access the meaningful parts; that is, we are able to focus attention on a specific feature of an image [Pylyshyn 1973]. Nested symbolic arrays capture these properties by representing images at various levels of abstraction (as suggested by the parts hierarchy of the deep representation). Each level of embedding in an array corresponds to a structural decomposition in the semantic network.

The spatial representation can be thought of as descriptive since it can be expressed as a propositional representation, where the predicates are spatial relationships and the arguments are concrete, imaginable, objects. Although information in the spatial representation can be expressed as propositions, the representations are not computationally equivalent; that is, the efficiency of the inference mechanisms is not the same. The spatial structure of images possesses properties not possessed by deductive propositional representations. These properties help avoid the combinatorial explosion of correct but trivial inferences that must be explicitly represented in a propositional system. Lindsay argues that spatial image representations (symbolic representations in our case) support non-deductive inference by a constraint satisfaction mechanism built into the processes that construct and access them [Lindsay 1988]. Consider, for example, the spatial representation of the image of the map of Europe. To retrieve the information on what countries are north of Germany, we need only search a portion of the symbolic array representing the map. Thus, although the information embodied in the spatial representation is only a subset of the data stored in the deep representation, the indexing of this information using an array data structure can make spatial reasoning more efficient.

Another advantage of symbolic arrays, with respect to propositional representations, concerns temporal reasoning. Any cognitive system, natural or artificial, should be able to deal with a dynamic environment in which a change in a single item of knowledge might have widespread effects. The problem of updating a system's representation of the state of the world to reflect the effects of actions is known as the *frame problem* [Raphael 1971]. Funt has previously shown that by representing the state of the world as a diagram, and actions in the world as corresponding actions in the diagram, helps to avoid many of the pitfalls of the frame problem [Funt 1980]. Representing an image as a symbolic array has similar advantages.

3.4 Primitive Functions

Approaches to knowledge representation are distinguished by the the operations performed on the representations. Thus, the effectiveness of our scheme can be partially measured by how well it facilitates the implementation of imagery related processes. In this subsection we review some of the primitive imagery functions that have been defined for the scheme. We also discuss how these functions provide the building blocks for more complex processes that mimic imagery capabilities.

As suggested earlier, array theory provides a metalanguage for specifying the representations and functions for imagery. In this theory, an array is a collection of zero or more items held at positions in a rectangular arrangement along zero or more axes. The mathematics of array theory provides for a multidimensional, hierarchical representation of the spatial and visual components of imagery which can make explicit features such as size, shape, symmetry, adjacency, etc. Arrays, in the theory, are nested structures; that is, arrays may occur as elements of other arrays. The concept of nesting, combined with the power of aggregating elements into a rectangular arrangement gives the theory much of its expressive power. There are several classes of predefined array functions in the theory, including logical operations, selectors, list operations, array transformers, etc. Since all of our representations (deep, spatial and visual) are stored as arrays, the primitive imagery functions can be defined in terms of array theory functions.

We consider the primitive functions for imagery in three classes corresponding to the three representations: deep, visual and spatial. Functions for the deep representation are just those of a frame language. These functions allow for construction and retrieval of frames. They also support retrieval of image properties through inheritance. Imagery functions for manipulating the visual representation include *rotate, translate* and *zoom*, which change the orientation, location or size of an image. Functions for retrieving *volume* and *shape* of an occupancy array representation of an image have also been define. Figure 1 presents a brief description of some of the operations that have been defined for spatial reasoning. In order to reason with images, we also provide functions that allow us to interpret the spatial representations in terms of propositions within a given domain.

4 Conclusions

This paper describes the concept of computational imagery, which treats imagery as a problem solving paradigm in AI. By proposing a knowledge representation scheme that attempts to capture the fundamental principles of mental imagery, we provide a foundation for implementing systems that rely on imagery-based reasoning.

Aside from related research in areas such as perception, the AI community has given little attention to the topic of imagery. Thus we rely on the relevant theories of cognitive psychology and neuroscience to provide initial guidance for our research. We are also driven by the need to apply the scheme to real world applications. The representation scheme is not intended to be a model of mental imagery; we do not

NAME	DESCRIPTION
retrieve	Reconstruct spatial image
put	Place an image component relative to another
find	Find location of component
delete	Delete image component
move	Move image component to new location
turn	Rotate image 90 degrees in specified direction
focus	Replace specified subimage with its spatial representation
unfocus	Return to original image
store	Stores current image in long-term memory
adjacent	Determine adjacent image components

Figure 1: Primitive functions for spatial imagery

claim that in human working-memory two "mind's eyes" exist that process visual and spatial representations identical to the ones that we have implemented. What we do suggest is that the internal image representations are informationally equivalent to representations involved in our scheme; that is, information in one representation is inferable from the other.

The knowledge representation scheme for computational imagery includes three image representations, each appropriate for a different kind of information processing. A set of primitive functions, which correspond to the fundamental processes involved in mental imagery, have been designed using the mathematics of array theory and implemented in the functional array language Nial. These functions provide the building blocks for more complex imagery related processes.

An approach to computational imagery should attempt to capture the five fundamental principles of mental imagery described in [Finke 1989]. In particular, the scheme was designed around the principle of *implicit encoding*, which states that imagery is used to extract information that was not explicitly stored in long-term memory. We retrieve information such as shape and size using the visual representation and information pertaining to the relative locations of objects in an image using the spatial representation for working memory. The principle of *perceptual equivalence* is captured by our assumption that perception and imagery share common representations. In fact the processes involved in transforming a visual representation to a spatial representation are just those of scene analysis - taking a raw uninterpreted image (visual representation) and identifying the subcomponents and their relative positions (spatial representation). The spatial representation captures the principle of *spatial equivalence*, since there is a correspondence between the arrangement of the parts of a symbolic array of an image, and the arrangement of the actual objects in the space. The principle of *structural equivalence* is preserved by the deep and the spatial representations, which capture the hierarchical organization of images. Furthermore, images in our representation scheme can be reorganized and reinterpreted. The scheme captures the functionality required of the principle of *transformational equivalence* by providing primitive array functions that can be used to manipulate both the visual and spatial representations of images.

Although it was not designed to be a psychological model of mental imagery, the scheme presented in the paper shares with these models some common characteristics. Similar to previous models, the scheme is based on an array theory of imagery.[1] According to such theories, mental images are array-like visual representations generated through perception, or from long-term memory, deep representations [Pinker 1988]. Several types of image arrays have previously been suggested. Kosslyn's model uses *2D* arrays for the implementation of the visual representation [Kosslyn 1980]. Pinker has suggested *2D* arrays combined with information about the *3D* shapes of objects in long-term memory files from which array patterns are generated [Pinker 1980]. A recent model, based on an array theory of imagery, describes how visual information can be represented within the computational framework of discrete symbolic representations in such a way that both mental images and symbolic thought processes can be explained [Chandrasekaran, Narayanan 1990].

A fundamental goal of our research is to use the representations and functions of computational imagery to develop knowledge-based systems that integrate the imagery problem solving paradigm with other reasoning techniques. One such system is an application to the problem of molecular scene analysis [Glasgow et al. 1991], which combines tools from the areas of protein crystallography and molecular-database analysis, through a framework of computational imagery. Other potential applications for

[1]The use of the term array theory here differs from our earlier usage. We base our representations and functions for computational imagery on an extended and more formal theory of arrays - the mathematics of nested, multi-dimensional collections of objects.

imagery-based systems include perception and medical imaging. Of special interest, are applications such as motion planning and game playing, which combine spatial and temporal reasoning.

When questioned on the most urgent unresolved difficulties in AI research, Aaron Sloman replied [Sloman 1985]:

> *"I believe that when we know how to represent shapes, spatial structures and spatial relationships, many other areas of AI will benefit, since spatial analogies and spatial modes of reasoning are so pervasive."*

Experimental results suggest that people use mental imagery for spatial reasoning. Thus, by facilitating an efficient implementation of the processes involved in mental imagery, computational imagery provides a basis for addressing the difficulties suggested by Sloman and developing AI systems that rely on representing, retrieving and reasoning about properties of images.

BIBLIOGRAPHY

Chandrasekaran, B. & Narayanan, N. H. (1990). Integrating imagery and visual representations. *Proceedings of the 12th Annual Conference of the Cognitive Science Society.* Lawrence Erlbaum Associates, 670-677.

Collins, A.M. & Quillian, M.P. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior,* 8, 240-247.

Finke, R.A. (1989). *Principles of Mental Imagery.* MIT press.

Finke, R.A., Pinker, S. & Farah, M., J. (1989). Reinterpreting visual patterns in mental imagery. *Cognitive Science,* 13, 51-78.

Funt B.V. (1980). Problem-solving with diagrammatic representations. *Artificial Intelligence,* 13, 201-230.

Glasgow, J.I. (1990). Artificial intelligence and imagery. *Proceedings of the Second IEEE Conference on Tools for AI.*

Glasgow, J.I., Fortier S. & Allen F.H. (1991). Crystal and molecular structure determination through imagery, In Hunter (Ed.), *Artificial Intelligence and Molecular Biology,* AAAI Press, in Press. Also appears in *Proceedings of the Seventh IEEE Conference on Artificial Intelligence Applications,* 98-105.

Jenkins, M.A. & Glasgow, J.I. (1989). A logical basis for nested array data structures. *Computer Languages,* 14, 35-49.

Jenkins, M.A., Glasgow, J.I. & McCrosky, C. (1986). Programming styles in Nial. *IEEE Software,* 86, 46-55.

Kosslyn, S.M. (1980). *Image and Mind.* Harvard University Press.

Kosslyn, S.M. (1987). Seeing and imagining in the cerebral hemispheres: a computational approach. *Psychological Review,* 94, 148-175.

Lindsay, R.K. (1988). Images and inference. *Cognition,* 29, 229-250.

Mishkin, J. Ungerleider, L.G. & Macko, K.A. (1983). Object vision and spatial vision: two cortical pathways, *Trends in Neuroscience,* 6, 414-417.

More, T. (1979). The nested rectangular array as a model of data. From proc. APL79, *APL Quote Quad,* 9.

Papadias, D. and Glasgow, J.I. (1991). A knowledge representation scheme for computational imagery. *Proceedings of the 13th Annual Conference of the Cognitive Science Society.* Lawrence Erlbaum Associates.

Pinker, S. (1980). Mental imagery and the third dimension. *Journal of Experimental Psychology: General,* 109, 354-371.

Pinker, S. (1988). A computational theory of the mental imagery medium. In Denis, Engelkamp & Richardson (Eds), *Cognitive and Neuropsychological Approaches to Mental Imagery.* Martinus Nijhorff Publishers, 17-36.

Pylyshyn, Z.W. (1973). What the mind's eye tells the mind's brain: a critique of mental imagery. *Psychological Bulletin,* 80, 1-24.

Pylyshyn, Z.W. (1981). The imagery debate: analogue media versus tacit knowledge. *Psychological Review,* 88, 16-45.

Raphael, B. (1971). The frame problem in problem-solving systems. In Findler and Meltzer (Eds.), *Artificial Intelligence and Heuristic Programming,* Edinburgh University Press, 159-169.

Sloman, A. (1985). In Bobrow & Hayes (Eds.), *Artificial Intelligence,* 25, 386.