COGNITIVE SCIENCE **16**, 355–394 (1992)

Computational Imagery

JANICE GLASGOW AND DIMITRI PAPADIAS

Queen's University

After many years of neglect, the topic of mental imagery has recently emerged as an active area of research and debate in the cognitive science community. This article proposes a concept of computational imagery, which has potential applications to problems whose solutions by humans involve the use of mental imagery. Computational imagery can be defined as the ability to represent, retrieve, and reason about spatial and visual information not explicitly stored in long-term memory.

The article proposes a knowledge representation scheme for computational imagery that incorporates three representations: a long-term memory, descriptive representation and two working-memory representations, corresponding to the distinct visual and spatial components of mental imagery. The three representations, and a set of primitive functions, are specified using a formal theory of arrays and implemented in the array-based language Nial. Although results of studies in mental imagery provide initial motivation for the representations and functionality of the scheme, our ultimate concerns are expressive power, inferential adequacy, and efficiency.

Numerous psychological studies have been carried out and several, often conflicting, models of mental imagery have been proposed. This article does not present another computational model for mental imagery, but instead treats imagery as a problem-solving paradigm in artificial intelligence (AI). We propose a concept of computational imagery, 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

The research described in this article was supported by grants from the Intelligent Robotics and Information Systems Federal Network Center of Excellence, the Information Technology Research Center of Ontario, and the Natural Science and Engineering Research Council of Canada. We would like to acknowledge crystallographers Suzanne Fortier and Frank Allen for their collaboration and contributions in the area of molecular scene analysis.

Correspondence and requests for reprints should be sent to Janice Glasgow, Department of Computing and Information Science, Queens University, Kingston, Ontario K7L 3N6, Ontario.

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, and so on. In particular, we are concerned with the reconstruction of image representations 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 to 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 hierarchically organized, 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 relevant spatial properties.
- The *visual representation* depicts the space occupied by an image as an occupancy array. It can be used to retrieve information such as shape, relative distance, and relative size.

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 (More, 1981). 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, Glasgow, & McCrosky, 1986).

The knowledge representation scheme for computational imagery provides a basis for implementing programs that involve reconstructing and reasoning with image representations. One such system, currently under investigation, is a knowledge-based system for molecular scene analysis.

Some of the concepts presented in this article will be illustrated with examples from that application area.

Research in computational imagery has three primary goals: a cognitive science goal, an AI goal, and an applications goal. The *cognitive science goal* addresses the need for computational models for theories of cognition. We describe a precise, explicit language for specifying, implementing, and testing alternative, and possibly conflicting, theories of cognition. The *AI goal* involves the development of a knowledge representation scheme for visual and spatial reasoning with images. Finally, the *applications goal* involves incorporating the knowledge representation scheme for computational imagery into the development of programs for solving real-world problems.

The article begins with an overview of previous research in mental imagery, which serves as a motivation for the representations and processes for computational imagery. This is followed by a detailed description of the deep, visual, and spatial representations for imagery, and the primitive functions that can be applied to them. It concludes with a summary of the major contributions of computational imagery to the fields of cognitive science, AI, and knowledge-based systems development, and a discussion of the relationship between our scheme and previous research in the area.

MENTAL IMAGERY

In vision research, an image is typically described as a projection of a visual scene of the back of the retina. However, in theories of mental imagery, the term "image" refers to an internal representation used by the human information-processing system to retrieve information from memory.

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, Pinker, & Farah, 1989). Others debate whether or not images play any causal role in the brain's information processing (Block, 1981). According to Farah (1988a), in depictive theories the recall of visual objects consists of the top-down activation of perceptual representation, but in descriptive theories visual recall is carried out using representations that are distinct from those in vision, even when it is accompanied by the phenomenology of "seeing with the mind's eye." Further discussions on the imagery debate can be found in various sources (e.g., Anderson, 1978; Block, 1981; Kosslyn & Pomerantz, 1977).

This article does not attempt to debate the issues involved in mental imagery, but to describe effective computational techniques for storing and manipulating image representations. To accomplish this, however, requires an understanding of the broad properties of representations and processes involved in mental imagery.

Research Findings in Mental Imagery

Many psychological and physiological studies have been carried out in an attempt to demystify the nature of mental imagery. Of particular interest to our research are studies that support the existence of multiple image representations and describe the functionality of mental imagery processes. In this section we overview relevant results from such studies, and based on these results, propose some important properties of mental imagery, which we use to motivate our representation scheme for computational imagery.

Several experiments provide support for the existence of a visual memory, distinct from verbal memory, in which recognition of verbal material is inferior. Paivio's (1975) dual-code theory suggests that there is a distinction between verbal and imagery processing. This theory leaves the exact nature of mental images unspecified, but postulates two interconnected memory systems—verbal and imaginal—operating in parallel. The two systems can be independently accessed by relevant stimuli but they are interconnected in the sense that nonverbal information can be transformed into verbal and vice versa. Furthermore, it has been indicated that visual memory may be superior in recall (Standing, 1973).

The issue of visual memory is an important one for computational imagery. What it implies to us is the need for separate descriptive and depictive representations. This is reinforced by the experiments carried out by Kosslyn (1980) and his colleagues, who concluded that images preserve the spatial relationships, relative sizes, and relative distances of real physical objects. Pinker (1988) suggested that image scanning can be performed in two- and three-dimensional space, providing support for Kosslyn's proposal that mental images capture the spatial characteristics of an actual display. Pinker also indicated that images can be accessed using either an object-centered or a world-centered coordinate system.

A series of experiments suggest that mental images are not only visual and spatial in nature, but also structurally organized in patterns, that is, they have a hierarchical organization in which subimages can occur as elements in more complex images (Reed, 1974). Some researchers propose that under certain conditions images can be reinterpreted: They can be reconstructed in ways that were not initially anticipated (Finke et al., 1989). Experiments also support the claim that creative synthesis is performed by composing mental images to make creative discoveries (Finke & Slayton, 1988 (Finke & Slayton, 1988).

The relationship between imagery and perception was considered by Brooks (1968), who demonstrated that spatial visualization can interfere with perception. Farah (1988a) also suggested that mental images are visual representations in the sense that they share similar representations to those used in vision, but noticed that this conclusion does not imply that image representations are depictive because both imagery and perception might be descriptive. Farah argued, from different evidence, however, that they are in fact spatial.

Findings, provided by the study of patients with visual impairments, point toward distinct visual and spatial components of mental imagery. Mishkin, Ungerleider, and Macko (1983) showed that there are two distinct cortical visual systems. Their research indicated that the temporal cortex is involved in recognizing *what* objects are, whereas the parietal cortex determines *where* they are located. Further studies have verified that there exists a class of patients who often have trouble localizing an object in the visual field, although their ability to recognize the object is unimpaired (De Renzi, 1982). Other patients show the opposite patterns of visual abilities: They cannot recognize visually presented objects, although they can localize them in space (Bauer & Rubens, 1985). Such patients are able to recognize objects by touch or by characteristic sounds. It has also been suggested that the preserved and impaired aspects of vision in these patients are similarly preserved or impaired in imagery (D. Levine, Warach, & Farah, 1985). In experimental studies, subjects with object identification problems were unable to draw or describe familiar objects despite being able to draw and describe in detail the locations of cities on a map, furniture in a house, and landmarks in a city. Patients with localization problems were unable to describe relative locations, such as cities on a map, although they could describe from memory the appearance of a variety of objects. Such findings have been interpreted by some researchers (e.g., Kosslyn, 1987) as suggesting two distinct components of mental imagery, the spatial and the visual, where the spatial component preserves information about the relative positions of the meaningful parts of a scene and the visual component preserves information about how (e.g., shape, size) a meaningful part of a scene looks.

Although there are varying strategies for retrieving spatial information and solving problems concerning spatial relations, research has suggested that humans typically use mental imagery for spatial reasoning (Farah, 1988b). Experimental results also support an isomorphism between physical and imaged transformations (Shepard & Cooper, 1982). A premise of Kritchevsky (1988) is that behavior can be divided into spatial and nonspatial components. For example, determining the color of an object is a nonspatial behavior, whereas determining relative positions of objects is a spatial behavior. Kritchevsky assumed that the spatial component of behavior is understood in terms of elementary spatial functions. Furthermore, these functions are independent of any particular sensory modality (Ratcliff, 1982).

Although individually the results described previously do not imply a particular approach to computational imagery, collectively they infer several properties that we wish to capture in our approach. Most importantly, an image may be depicted and reasoned with visually or spatially, where a visual representation encodes what the image looks like and the spatial representation encodes relative location of objects within an image. As well, images are inherently three-dimensional and hierarchically organized. This implies that computational routines must be developed that can decompose, reconstruct, and reinterpret image representations. Results from studies comparing imagery and vision imply that the representations and processes of imagery may be related to those of high-level vision. Thus, we should also consider the representations and functionality of object recognition when defining computational imagery. Finally, we must be able to consider an image from either an object-centered or a viewer-centered perspective.

The numerous experiments that have been carried out in mental imagery not only suggest properties for the representation scheme, but also support the premise that mental imagery is used extensively to reason about real-world problems. Thus, computational imagery is an important topic to investigate in relation to AI problem solving.

The subjective nature of mental imagery has made it a difficult topic to study experimentally. Qualities like clarity, blurring, and vividness of images are not directly observable and may differ from one person to another. Furthermore, it has been argued by some researchers that it is impossible to resolve the imagery debate experimentally because depictive and descriptive representations do not have distinct properties from which behavioral consequences can be predicted (Anderson, 1978). As a result, several alternative accounts have been proposed to explain the findings mentioned previously. The most important of these are tacit knowledge, experimenter bias, eye movements, and task-induced characteristics (Intons-Peterson, 1983). These difficulties involved in experimental studies emphasize the need for computer models for mental imagery. Although the knowledge representation scheme for computational imagery is not meant to model a particular theory of imagery, it does provide the tools for specifying, testing, and formally analyzing a variety of theories, and thus can contribute to resolving the imagery debate.

Theories and Principles of Mental Imagery

Pylyshyn (1981), a forceful proponent of the descriptive view, argued 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 disputed the idea that mental images are stored in a raw uninterpreted form resembling mental photographs, and argued for an abstract format of representation called propositional code. Kosslyn's (1980) model of mental imagery 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 referred to as the "mind's eye," serves as an interface between the visual representations and the underlying data structures, which may be decidedly nonpictorial in form. Hinton (1979) disputed the picture metaphor for imagery and claimed that images are more like generated constructions. In this approach, as in Marr and Nishihara's (1978) 3D model, complex images can be represented as a hierarchy of parts.

Finke (1989) took a different approach to the imagery debate. Instead of proposing a model, Finke defined five "unifying principles" of mental imagery:

- The principle of *implicit encoding* states that imagery is particularly useful for retrieving information about physical properties of objects and relations among objects whenever this information was not previously, explicitly encoded.
- The principle of *perceptual equivalence* states that similar mechanisms in the visual system are activated when objects or events are imagined, as when the same objects or events are actually perceived.
- The principle of *spatial equivalence* states that the spatial relations between objects are preserved, although sometimes distorted, in mental images.
- The principle of *structural equivalence* states that the structure of images corresponds to that of perceived objects, in the sense that the structure is coherent, well organized, and can be reinterpreted.
- The principle of *transformational equivalence* states that imagined and physical transformations exhibit similar dynamic characteristics and follow the same laws of motion.

These principles provide a basis for evaluating the representations and functions for computational imagery; in the development of our scheme we have attempted to address each of the underlying principles for mental imagery.

Stages of Image Representations

The hypothesis of multiple representations for mental imagery can explain several experimental results that cannot be explained independently by either a propositional, a spatial, or a visual representation. For instance, after a series of experiments, Atwood (1971) concluded that memory for high-image phrases is disrupted if followed by a task requiring the subject to process a visually presented digit in contrast to abstract phrases. Although other researchers found difficluty in replicating Atwood's experiments, Jannsen (1976) succeeded consistently over several experiments and claimed that other failures stemmed from using an interfering task that is spatial rather than visual. Baddeley and Lieberman (1980) interpreted these results as pointing towards distinct visual and spatial components of mental imagery.

When images are retrieved, it is possible to recall information about which objects constitute a scene and their spatial relationships with other objects without remembering what the object looks like. Furthermore, we are able to recognize objects independent of any context. Distinct spatial and visual components for imagery can explain such phenomena, where the spatial component can be considered as an index that connects visual images to create a scene.

Intuitively, we can distinguish between visual and spatial representations by considering the type of information we wish to retrieve. Consider, for example, answering the following questions: *How many windows are there in your home? What city is farther north, Seattle or Montreal? What objects are sitting on top of your desk? Who was sitting beside Mary in class?* These questions can typically be answered without constructing an explicit visual image, that is, you could possibly recall that John was sitting beside Mary without knowing what John looked like or what clothes he was wearing. Each of these questions does rely on knowing the relative locations of objects within a recalled image, information that is embodied in a spatial representation. Now consider questions such as: *What is the shape of your dog's ears? What does a particular image look like if you rotate it ninety degrees? What is larger, a rabbit or a racoon? Is Montreal or Toronto closer to Ottawa?* To answer these questions you may need to reconstruct a representation that preserves information such as size, shape, or relative distance, information that is embodied in a visual representation.

From the computational point of view, a single representational system cannot always effectively express all the knowledge about a given domain; different representational formalisms are useful for different computational tasks (Sloman, 1985). In perceptual systems, for instance, multiple representations have been proposed to derive cognitively useful representations from a visual scene. For computational imagery, we propose three stages of image representation, each appropriate for a different type of information processing (Papadias & Glasgow, 1991). The deep representation stores structured, descriptive information in terms of a semantic network, long-term memory model. The working-memory representations (spatial and visual) are consciously experienced and generated as symbolic and occupancy arrays, as needed, using information stored in the deep representation. Details about the computational advantages of each of the image representations involved in the scheme will be presented in the following section.

KNOWLEDGE REPRESENTATION SCHEME

Research in AI has long been concerned with the problem of knowledge representation. AI programs rely on the ability to store descriptions of a particular domain and formally manipulate these descriptions to derive new knowledge. Traditional approaches to knowledge representation include logic representations, which denote the objects and relations in the world in terms of axioms, and structural knowledge representation schemes, which denote concepts and relations in terms of structural hierarchies.

In addition to general schemes, there exist specialized schemes concerned with the representation of the visual representation of images. In discrimination trees, objects are sorted by discriminating on their coordinates, as well as other quantitative and qualitative discriminators (McDermott & Davis, 1984). A simple way of describing volume or shape is with occupancy arrays, where cells of the array denote objects filling space. For computer vision applications, an occupancy array is often called a gray-level description, because the value of the cells encode the intensity of light on a gray scale from white to black. For our molecular scene analysis application, we use three-dimensional occupany arrays that correspond to electron density maps resulting from X-ray diffraction experiments. The values of the cells in such maps correspond to the electron density in a unit cell of a crystal.

According to Biederman (1987), the visual representation for objects can be constructed as a spatial organization of simple primitive volumes, called geons. Other researchers have proposed alternative primitive volumes, like generalized cones, spheres, and so forth. A major contribution in representational formalisms for images is the progression of primal sketch, 2½D sketch, and 3D sketch (Marr & Nishihara, 1978). The primal sketch represents intensity changes in a 2D image. The 2½D sketch represents orientation and depth of surface from a particular viewer perspective. Finally, the 3D sketch represents object-centered spatial organization.

The representation schemes discussed before are not suggested as structures for representing human knowledge and do not necessarily commit to addressing questions about mental processes. Whereas many AI researchers believe that the best way to make true thinking machines is by getting computers to imitate the way the human brain works (Israel, 1987), research in knowledge representation often is more concerned with expressiveness and efficiency, rather than explanatory and predictive power. Thus, although our knowledge representation scheme attempts to preserve the most relevant properties of imagery, whenever possible we try to overcome the limitations of the human information-processing system. For example, theories of divided attention argue that attention can be concentrated on, at most, a few mental processes at a time. Our proposed scheme has the capability of relatively unrestrictive parallel processing of spatial images. Furthermore, although the resolution of mental images is limited by the capabilities of the human mind, in the knowledge representation scheme the resolution restrictions are imposed by the implementation architecture.

A theory of arrays provides a formalism for the representations and functions involved in computational imagery. Array theory (More, 1981) is the mathematics of nested, rectangularly arranged collections of data objects. Similar to set theory, array theory is concerned with the concepts of nesting, aggregation, and membership. Array theory is also concerned with the concept of data objects having a spatial position relative to other objects in a collection. Thus, it provides for a multidimensional, hierarchical representation of images, in which spatial relations are made explicit.

We consider 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, whereas 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. As illustrated in Figure 1, 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 describe each of the representations in detail and discuss the primitive functions that operate on them. First, though, we overview the theory or arrays that provide the basis for describing and implementing the representations and functions for computational imagery.

Array Theory

Results of empirical studies suggest that images may be organized using both a hierarchical and a spatial structure. Components of an image may be grouped into features and stored based on their topological relations, such as adjacency or containment, or their spatial relations, such as *above, beside, north-of,* and so on. Because of the relevance of storing and reasoning about such properties of an image, we base the development of the knowledge representation scheme for computational imagery on a theory of arrays. This mathematical theory allows for a multidimensional, hierarchical representation of images in which spatial relations are made explicit. Furthermore, functions can be defined in array theory for constructing, manipulating, and retrieving information from images represented as arrays. For example, functions that compose, translate, juxtapose, and compare images have been defined within the theory.

The development of array theory was motivated by efforts to extend the data structures of APL and has been influenced by the search for total operations that satisfy universal equations (More, 1981). In this theory, an array is a collection of zero or more items held at positions in a rectangular arrangement along multiple axes. Rectangular arrangement is the concept of data objects having a position relative to other objects in the collection.

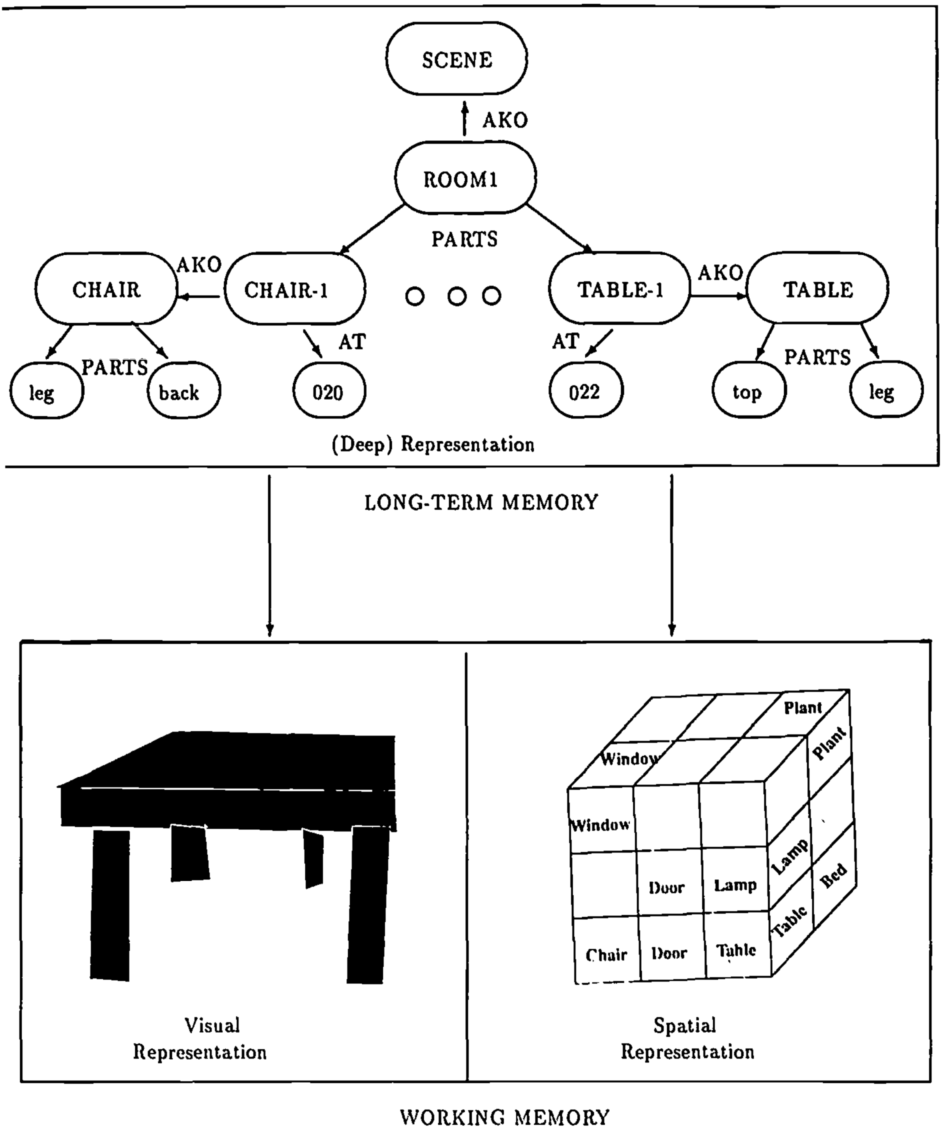


Figure 1. Representations for computational imagery

The interpretation of this structure can be illustrated using nested, box diagrams. Consider the array diagram in Figure 2. In this array the pair formed from 7 and 9 is an array nested within the larger array. Nesting is the concept of having the objects of a collection be collections themselves. This is an important concept in array theory because it is the power of aggregating arbitrary elements in an array that gives the theory much of its expensive

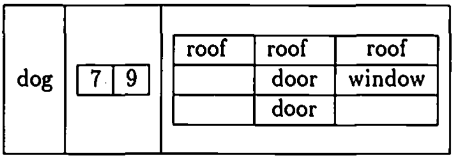


Figure 2. Example of nested array diagram

power. The third element of the array is a symbolic array, which denotes an image of a house containing three parts. The indexing of the array allows us to make explicit such properties as *above(roof,door)* and *left-of(door, window)* in a notation that is both compact and accessible.

Array theory has provided a formal basis for the development of the *N*ested *I*nteractive *A*rray *L*anguage, *Nial*. This multiparadigm programming language combines concepts from APL, Lisp, and FP with conventional control mechanisms (Jenkins et al., 1986). The primitive functions of array theory have all been implemented in Q'Nial (Jenkins & Jenkins, 1985), a commercially available, portable interpreter of Nial developed at Queen's University.

Operations in array theory are functions that map arrays to arrays. A large collection of total, primitive operations are described for the theory. They are chosen to express fundamental properties of arrays. Nial extends array theory by providing several syntactic forms that describe operations, including composition, partial evaluation of a left argument, and a lambda form. Array theory also contains second-order functions called transformers that map operations to operations.

It has previously been shown that the syntactic constructs of array theory facilitate both sequential and parallel computations (Glasgow, Jenkins, McCrosky, & Meijer, 1989). This is an important feature when considering computational imagery as a basis for specifying cognitive processes, which themselves may be sequential or parallel. The potential parallelism in array theory comes from three sources: inherent parallelism in the primitive operations, parallelism expressed by syntactic constructs, and parallelism in operation application controlled by primitive transformers. The potential parallelism of the primitive operations results from treating an entire array as a single value; each array takes an array as a single argument and returns an array as its result. Array theory includes transformers that allow expression of the parallel application of an operation to subparts of an array.

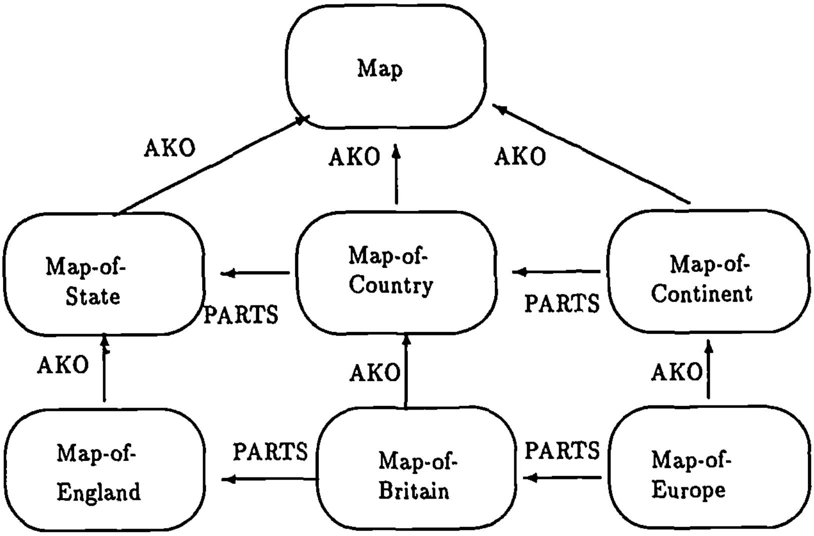
The software development associated with AI problem solving in general, and with computational imagery in particular, differs from traditional computer applications. AI problems are solved at the conceptual level, rather than a detailed implementation level. Thus, much of the programming effort is spent on understanding how to represent and manipulate the knowledge associated with a particular problem, or class of problems. This imposes certain features on a programming language, including interactive program development, operations for symbolic computation, dynamically created data structures, and easy encoding of search algorithms. Although Lisp and Prolog both address capabilities such as these, they provide very different and complementary approaches to problem solving. The language Nial is an attempt to find an approach to programming that combines the logic and functional paradigms of Prolog and Lisp (Glasgow & Browse, 1985, Jenkins et al., 1986). It has been demonstrated that array theory and Nial can provide a foundation for logic programming (Glasgow, Jenkins, Blevis, & Feret, in press), as well as other descriptive knowledge representation techniques (Jenkins et al., 1988). These techniques have been implemented and tested on a variety of knowledge-based applications.

Deep (Long-Term Memory) Representation

The deep representation for computational imagery is used for the long-term storage of images. Earlier work has suggested that there exists a separate long-term memory model that encodes visual information descriptively (Kosslyn, 1980; Pinker, 1984). This encoding can then be used to generate depictive representations in working memory. As pointed out in Marschark, Richman, Yuille, and Hunt (1987), most of the research in vision and imagery has focused on the format of the on-line conscious representations, excluding long-term storage considerations. 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 because 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 (Minsky, 1975), in which each frame contains salient information about an image or class of images. This information includes propositional and procedural knowledge. 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 can be characterized as nonmonotonic because default information (stored in specific slots, or inherited from more generic frames) is superseded as new information is added to a frame.

A frame corresponding to the image of a map of Europe and part of the semantic network for a map domain is illustrated in Figure 3. Each node in the network corresponds to an individual frame and the links describe the relationships among frames. The AKO slot in the frame of the map of



a) Semantic network representation

FRAME	Map-of-Europe
AKO	Map-of-Continent
PARTS	Sweden (0 4) Britain (1 0) ...
POPULATION	'find-population'
...	...

b) Frame representation

Figure 3. Example of deep representation

Europe denotes that the frame is an instance of the concept "Map-of-Continent." The PARTS slot contains the meaningful parts that compose the map, along with an index value that specifies their relative locations. The POPULATION slot contains a call to a procedure that calculates the population of Europe, given the populations of the countries. As well, the frame could incorporate several other slots, including ones used for the generation of the spatial and visual representations.

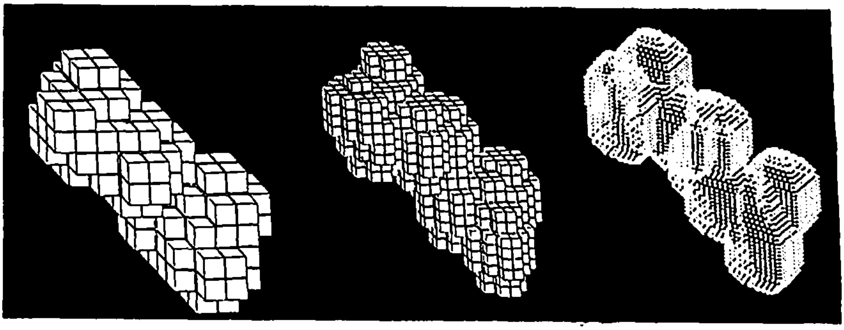
For the molecular scene analysis application, the frame hierarchy is more complex than the simple map example. The structure of a protein is described in terms of a crystal, which consists of a regular three-dimensional arrangement of identical building blocks. The structural motif for a protein crystal can be described in terms of aggregate (complex or quaternary), three-dimensional structures. Similarly, tertiary structures can be decomposed into secondary structures, and so on. Each level in this decomposition hierarchy corresponds to a conceptual frame denoting a molecular fragment at a meaningful level of abstraction. If we consider a fully determined crystal as a molecular scene, there exist databases containing over 90,000 images of small molecules and over 600 images of protein structures (Allen, Bergerhoff, & Sievers, 1987). These databases include the three-dimensional geometry of the molecular scenes that forms a basis for our long-term memory model for molecular images.

Semantic networks and frames have previously been suggested as representations for images in vision research. One example of this deals with the interpretation of natural scenes (M. Levine, 1978). In Levine's system, the spatial relations are represented as arcs such as *left-of, above,* or *behind.* A classic example of the use of semantic networks is the work of Winston (1975) on structural descriptions. In that study on scene understanding, common structures, such as arches and pedestals, are represented in terms of their decomposition into parts and a description of the spatial relations among the parts. Although this approach may be useful for some applications, we argue later that explicitly representing spatial relations in terms of an indexed array provides increased computational efficiency for spatial reasoning.

Our implementation of the deep representation has several attractive properties. First, it provides a natural way to represent knowledge because 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 is assumed that a property is stored at the most general level possible (highest level in the conceptual hierarchy) and is shared by more specific levels, thus providing a large saving in space over propositional or database formulations of property relations. The deep representation also incorporates the psychological concept of semantic networks in an implementation that provides features such as procedural attachment. The nonmonotonic feature of the frame allows for reasoning with incomplete information; default information can be stored in conceptual frames and inherited and used for depicting or reasoning about subconcepts or instances of images. Despite its attractive properties, however, 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.

Working-Memory Representations

Mental images are not constantly experienced. When an image is needed, it is generated on the basis of stored information. Thus, unlike the deep representation, which is used for the permanent storage of information, the



a) 1.0A b) 0.5A c) 0.15A

Figure 4. Example of occupancy arrays for visual representations

working-memory representations of an image exist only during the time that the image is active, that is, when visual or spatial information processing is taking place.

The distinct working-memory representations were initially motivated by results of cognitive studies that suggest distinct components in mental imagery (Kosslyn, 1987). More importantly, separate visual and spatial representations provide increased efficiency in information retrieval. The visual representation is stored in a format that allows for analysis and retrieval of such information as shape and relative distance. Because the spatial representation makes explicit the important features and structural relationships in an image while discarding irrelevant features such as shape and size, it provides a more compact and efficient depiction for accessing spatial and topological properties.

Visual Representation. 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 (1980) skeletal image, this representation is depictive and incorporates geometric information. Unlike Kosslyn's approach, we assume that the visual representation can be three-dimensional 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, and surface orientation about this region. Objects are depicted in the arrays by patterns of filled cells isomorphic in surface area to the objects. Figure 4 illustrates depictions of three-dimensional occupany arrays corresponding to a molecular fragment at varying levels of resolution. These arrays were constructed using geometric coordinates and radii of the atomic components of the molecule.

Representing occupancy arrays explicitly in long-term memory can be a costly approach. As a result, other approaches to storing or generating this information (like generalized shapes) have been developed. Such approaches can be incorporated into an application of the scheme for computational imagery.

Spatial Representation. A primary characteristic of a good formalism for knowledge representation is that it makes relevant properties explicit. Although 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, corresponding to the mental maps created by humans. 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 preserves the spatial and topological relationships of the image features, but not necessarily relative sizes or distances. The arrays can be interpreted in different ways depending on the application. If, for example, we use the scheme to reason about 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,* and so on. For molecular scene analysis we are more concerned with properties such as *symmetry* and *adjacency* (bonding), which are made explicit by a symbolic array. The spatial representation can also denote nonspatial dimensions. For example, the symbolic array could be used to 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. For example, in Figure 3 the PARTS slot contains the indices needed to reconstruct the spatial representation for a simplified map of Europe. Figure 5 illustrates this symbolic array. Note that some parts occupy more than one element in an array (e.g., Italy, France). This is necessary to capture all the spatial relationships of the parts of an image. We may also wish to denote more complex relations, such as one object being "inside" another. This is illustrated in Figure 6, which displays a spatial image of a glass containing water.

According to Pylyshyn (1973), images are not raw, uninterpreted, mental pictures, but are organized into meaningful parts that are remembered in terms of their spatial relations. Furthermore, we can access the meaningful parts, that is, we are able to focus attention on a specific feature of an

				Sweden	
Britain			Denmark		
		Holland	Germany	Germany	
		Belgium			
	France	France	Italy	Yugoslavia	Yugoslavia
Portugal	Spain		Italy		Greece

Figure 5. Example of symbolic array for spatial representation

glass	water	glass
glass	glass	glass

Figure 6. Symbolic array depiction of inside relation

image. Nested symbolic arrays capture these properties by representing images at various levels of abstraction as prescribed by the PART hierarchy of the deep representation; each level of embedding in an array corresponds to a level of structural decomposition in the frame hierarchy. For instance, focusing attention on Britain in the array of Figure 5 would result in a new array in which the symbol for Britain is replaced by its spatial representation (see Figure 7). This subimage is generated using the PARTS slot for the frame of Britain in the deep representation.

It has been suggested that people can reconstruct and reinterpret mental images (Finke, 1989). The proposed scheme also provides the capability to combine and reconstruct images, using special functions that operate on the symbolic array representations. For instance, we can combine a portion of the array of Figure 5 with a portion of the array that corresponds to the map of Africa and create a new array containing Mediterranean countries.

Recall that Pinker (1988) pointed out that images are represented and manipulated in three dimensions. Similar to the visual representation, a symbolic array can be two- or three-dimensional, depending on the application. In the domain of molecular scenes, fragments of molecules are represented as three-dimensional symbolic arrays at varying levels of abstraction, corresponding to the level of decomposition in the frame hierarchy. For example, a protein can be represented as a three-dimensional array of symbols denoting high-level structures, which can be decomposed into nested

<table>
<tr><td></td><td></td><td></td><td></td><td>Sweden</td><td></td></tr>
<tr><td><table><tr><td></td><td>Scotland</td></tr><tr><td>Wales</td><td>England</td></tr></table></td><td></td><td></td><td>Denmark</td><td></td><td></td></tr>
<tr><td></td><td></td><td>Holland</td><td>Germany</td><td>Germany</td><td></td></tr>
<tr><td></td><td></td><td>Belgium</td><td></td><td></td><td></td></tr>
<tr><td></td><td>France</td><td>France</td><td>Italy</td><td>Yugoslavia</td><td>Yugoslavia</td></tr>
<tr><td>Portugal</td><td>Spain</td><td></td><td>Italy</td><td></td><td>Greece</td></tr>
</table>

Figure 7. Embedded symbolic array representation

arrays of symbols denoting progressively more detailed substructures. Because of the size and complexity of molecular structures, it is essential to be able to reason at multiple levels of abstraction when analyzing a particular molecular scene. Figure 8 depicts a three-dimensional image of a fragment of a protein secondary structure, and an embedded amino acid residue substructure containing symbols denoting atoms. Bonding at the residue and atomic level is made explicit through structural adjacency in the representation.

For image recognition and classification, it is necessary to pick out characteristic properties and ignore irrelevant variations. One approach to image classification is on the basis of shape. Although the visual representation provides one approach to shape determination, the spatial representation allows for a hierarchical, topological representation for shape. This approach is particularly useful in applications where images are subject to a large number of transformations. For example, a human body can be configured many ways depending on the positions of the arms, legs, and so forth. Although it is impossible to store a separate representation for every possible configuration, it is possible to represent a body using a symbolic array that makes explicit the parts of the body and the relations among parts that remain constant under allowable transformations. Figure 9 illustrates such a spatial representation. Combined with a primitive shape descriptor (such as generalized cylinder), the spatial representation provides for multidimensional shape descriptors as proposed by Marr (1982).

The spatial representation can be thought of as descriptive because 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

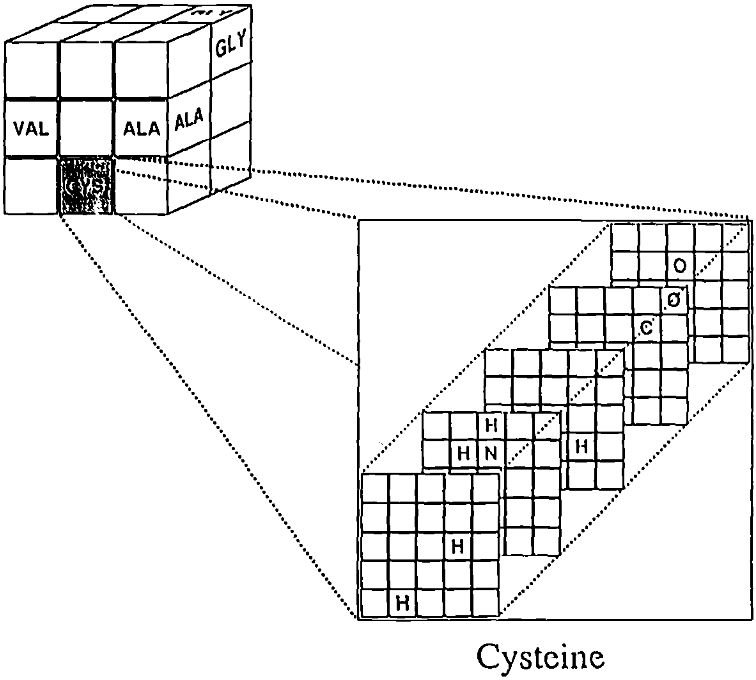


Figure 8. Symbolic array of molecular fragment

propositions, the representations are not computationally equivalent, that is, the efficiency of the inference mechanisms is not the same. The spatial structure of images has properties not possessed by deductive propositional representations. As pointed out by Lindsay (1988, p. 231), these properties help avoid the "combinatorial explosion by correct but trivial inferences that must be explicitly represented in a propositional system." Lindsay also argued that the spatial image representations (symbolic representations in our case) support nondeductive inference using built-in constraints on the processes that construct and access them. Consider, for example, the spatial representation of the map to Europe. To retrieve the information about what countries are north of Germany, we need only search a small portion of the symbolic array. Alternatively, in a propositional approach, the spatial relations would be stored as axioms such as

north-of(Britain,Portugal), north-of(France,Spain), north-of(Holland,Belgium)...,

and general rules such as

$$north\text{-}of(X,Y) \wedge north\text{-}of(Y,Z) \rightarrow north\text{-}of(X,Z).$$