Title
A Knowledge Representation Scheme for Computational Imagery

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 13(0)

Authors
Papadias, Dimitri
Glasgow, Janice

Publication Date
1991

Peer reviewed

A Knowledge Representation Scheme for Computational Imagery

Dimitri Papadias Janice Glasgow *
Department of Computing and Information Science
Queen's University, Kingston, Canada
janice@qucis.queensu.ca

Abstract

After many years of neglect, the topic of mental imagery has recently emerged as an active area of debate. One aspect of this ongoing debate is whether an image is represented as a *description* or a *depiction* of its components. This paper is not so concerned with how mental images are stored, as with what machine representations will provide a basis for imagery as a problem solving paradigm in artificial intelligence. In fact, we argue that a knowledge representation scheme that combines the ability to reason about both descriptions and depictions of images best facilitates the efficient implementation of the processes involved in imagery.

1 Introduction

Numerous experimental and neuropsychological 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 (AI). Our thesis is that there exists a concept of *computational imagery*, which has potential applications to problems whose solutions by humans involve the use of mental imagery. In order to define the concept of computational imagery, we require a knowledge representation scheme. Although psychological theories are used as a guide to the properties of mental imagery that our scheme should preserve, 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.

The knowledge representation scheme for computational imagery includes three representations - the deep representation, the symbolic array representation and the surface representation - each appropriate for a different kind of processing. While the deep representation is used as a permanent store for information, the symbolic array and surface representations are only constructed when retrieving spatial information not explicitly stored in the deep representation.

2 Mental Imagery

Although no one seems to deny the existence of the phenomenon called imagery, there has been an ongoing debate about the structure and the function of imagery in human cognition. The imagery debate is primarily 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). The debate can also be expressed as whether images are epiphenomenal or not, where epiphenomena are effects that do not play any causal role in the brain's information processing (Block 1981).

This paper does not attempt to debate the issues involved in mental imagery, but to describe effective computational techniques for storing and manipulating imagery-related representations. To achieve this goal, though, we believe it is necessary to better understand the processes involved in mental imagery.

Results in the area of cognitive psychology have suggested several properties of mental imagery that the proposed knowledge representation scheme should attempt to capture. Finke summarizes these properties by proposing five "unifying principles" of mental imagery: the principles of implicit encoding, perceptual equivalence, spatial equivalence, transformational equivalence and structural equivalence (Finke 1989).

While individual experiments attempt to explain a limited number of imagery phenomena, theories are

*Support for this research was provided by the Institute for Robotics and Intelligent Systems Federal Center of Excellence of Canada and the Natural Science and Engineering Research Council of Canada.

more general and attempt to explain a large number of such phenomena. 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 a depictive theory, which claims that images are quasi-pictorial. According to Kosslyn's model, mental images are working memory, surface representations generated from long-term memory, deep representations. A set of procedures serves as an interface between the surface representations and the underlying data structures, which may be decidedly non-pictorial in form.

Interesting findings, provided by the study of patients with visual impairments, have been interpreted by some researchers, for example (Kosslyn 1987), as suggestive of two distinct components of mental imagery: the spatial and the visual. The *spatial* component of imagery preserves information about the relative positions of the meaningful pars of an image. The *visual* component preserves information about *how* (size, shape, etc.) an image looks.

We do not propose a computational model for imagery. Rather, we present a knowledge representation scheme that attempts to capture the functionality of mental imagery. Although the scheme is similar to Kosslyn's in some ways, it extends his approach by defining image representations that are *3D* and hierarchical. Our approach also considers two working memory representations, corresponding to the visual and spatial components of imagery.

3 Knowledge Representation

AI programs rely on the ability to store domain descriptions and formally manipulate these descriptions to derive new knowledge about the given domain. Traditional approaches to knowledge representation include logic representations, which denote the objects and relations in the world in terms of logic rules, and structural representations, which denote concepts and relations in terms of hierarchies. In addition to general knowledge representation schemes, there exist specialized schemes concerned with the representation of shape, volume and other spatial qualities of images. These include discrimination trees (McDermott & Davis 1984), quadtrees (Samet 1980) and simple primitive volumes (Biederman 1985). A major contribution in representational formalisms for visual images is the progression of primal sketch, *2-1/2D* sketch and *3D* sketch (Marr & Nishihara 1978).

Kosslyn, in a discussion of the implementation of his computational model for mental imagery, expresses frustration with existing computational tools as well as the need for a metadescription for theories of imagery:

> *There is a major problem with this approach however; the program will not actually run without numerous "kluges", numerous ad hoc manipulations required by the realities of working with a digital computer and a programming language like ALGOL or LISP... The ideal would be a precise, explicit language in which to specify the theory and how it maps into the program.* (Kosslyn 1980)

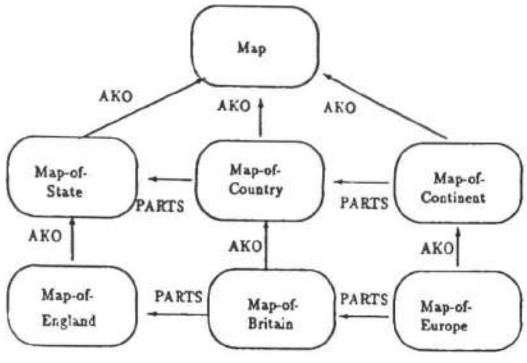
A formal theory of arrays provides such a metalanguage. Array theory is the mathematics of nested, rectangularly-arranged collections of data objects (More 1979)(Jenkins & Glasgow 1989). Similar to set theory, array theory is concerned with the concepts of nesting, aggregation and membership. Array theory is also concerned with the relative spatial positions of data objects in a collection.

Nial, an interactive programming language that combines concepts from LISP and APL, constitutes an implementation of array theory (Jenkins, Glasgow & McCrosky 1986). Thus, the language of array theory provides for a specification of the representations and functions for computational imagery as well as a mapping into Nial programs.

3.1 Deep Representation

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. In our view, the deep representation falls more within the limits of research in long-term memory than imagery. Thus, we base its implementation on the hierarchical network model of semantic memory (Collins & Quillian 1969). This network model is suitable for storing images since it can be used to denote the decomposition and classification hierarchies involved in imagery, as well as the features of concepts and instances of images.

There are two kinds of hierarchies in the network representation: the AKO (a kind of) and the PARTS. The AKO hierarchy provides property inheritance: an



a) Semantic network representation

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

b) Frame representation

Figure 1: Deep representation

image can inherit features from more generic images. The PARTS hierarchy is used to denote the structural decomposition of images into their meaningful components.

For the implementation of the deep representation in our scheme we use a frame language, in which each image frame contains (or inherits) all the salient information about the image. This information includes propositional knowledge, in addition to encodings that permit the reconstruction of the working-memory representations.

A frame corresponding to the image of a map of Europe and part of the semantic network for a geographic map domain is illustrated in Figure 1. Each node in the network corresponds to an individual frame and the links describe the structural and conceptual relationships between frames.

The frame implementation of the deep representation has several attractive properties. First it provides a natural way to represent knowledge since the hierarchical structure of an image is captured by the network structure. It also incorporates the concept of semantic networks in a computational implementation that provides features, such as multiple hierarchies and procedural attachment (ability to attach procedures to slot values). The non-monotonic feature of defaults in the frame structure corresponds to 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 computations. Certain tasks, such as determining the spatial relations between image components, surface pattern matching and image scanning, require information that is not explicitly stored in long-term memory. Thus, the construction of representations corresponding to the visual and spatial components of imagery are required to facilitate the efficiency of the scheme.

3.2 Working-Memory Representations

Unlike the deep representation, which is used for the permanent storage of information, the working-memory representations of an image exist only while the image is active (when visual or spatial information processing takes place).

3.2.1 Surface Representation

The surface representation corresponds to the visual component of imagery, and it can either be reconstructed from the underlying deep representation or generated from low level vision processes. Similar to Kosslyn's skeletal image (Kosslyn 1980), the surface representation is depictive and incorporates geometric information, including shape and volume. Unlike these representations, we assume that the surface representation can be *3D*. In such cases, we assume an object-centered description of an image.

For the current implementation of the surface representation we use occupancy arrays. An occupancy array consists of cells, each mapped to a local region of the visual field and representing information such as lightness, color, texture, edge orientation or depth about this region. Objects are depicted in the arrays by patterns of filled cells isomorphic in shape to the objects. Figure 2 illustrates a *2D* occupancy array corresponding to a map of Italy. Even at low resolution, this surface representation contains information about the shape and size of the country.

3.2.2 Symbolic Array 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

		1	1	1	1				
	1	1	1	1	1	1			
1	1	1	1	1	1	1			
			1	1	1				
			1	1	1				
				1	1	1			
				1	1	1			
					1	1	1		
					1	1	1		
						1	1		
					1	1	1		
				1	1			1	
			1	1	1				1

Figure 2: Occupancy array

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

Figure 3: Symbolic array representation

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

Figure 4: Embedded array representation

are best served by a descriptive representation that 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 elements of the array denote the meaningful components of an image. The symbolic array captures the spatial relationships of image components, but not necessarily relative sizes or distances. The interpretation of the spatial relations in a symbolic array is dependent on the given domain. If, for example, we use the scheme to represent 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.

Symbolic arrays can be generated using the frame's procedural attachment facility. The information to define this representation can either be stored explicitly, as in the example of Figure 1 where the PARTS slot contains the indices for the subimages, or derived from a geometric description of an image. Figure 3 illustrates the symbolic array corresponding to the image of a map of Europe. 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 image components. Also note that there is no attempt to capture size or distance information in this representation.

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. Furthermore, we can access the meaningful parts; we are able to focus attention on a specific feature of an image (Pylyshyn 1973). Nested symbolic arrays capture this property of mental imagery by representing images at various levels of abstraction (as suggested by the structural hierarchy of the deep representation). For instance, focusing attention on Britain in the array of Figure 3 would result in a new array in which the symbol for Britain is replaced by the symbolic array specifying its parts (see Figure 4). This subimage is generated from the frame representation for the image of Britain.

Similar to the surface representation, a symbolic array can be *2D* or *3D* depending on the application. The symbolic 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 symbolic representation can be expressed as propositions, the two representations are not computationally equivalent. The spatial structure of images possesses properties not found in 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 image of the map of Europe. Determining the countries north of Germany can be achieved by mentally scanning a small portion of the mental image of a map. Similarly, in the corresponding symbolic array we only need to search the subarray above Germany.

Another advantage of symbolic arrays, with respect to propositional representations, concerns the frame problem. 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 on many other items. Suppose that we change the position of a country in our map of Europe. Using propositions, we might need to change several relations involving the given country, and perhaps relations derived from the initial propositions. Using a symbolic array to store the map, we need only delete the country from its previous position and insert it in the new one. Since spatial relationships are determined functionally (not logically inferred) from image representations, we eliminate the problem of nonmonotonicity in domains involving spatial reasoning.

3.3 Primitive Functions for Imagery

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. Several primitive functions for imagery have been specified in array theory and implemented in Nial. These include functions for constructing images, transforming images and accessing images.

Images can be constructed in unique and creative ways by defining a symbolic array whose components correspond to existing frames in the deep representation. New images can also be created through low level perception techniques that transform an uninterpreted surface representation into a symbolic array. Operations that transform the symbolic and surface representations through rotation and translation have been implemented. As well, a focus function, which transforms a symbolic array by replacing a symbol with its corresponding symbolic array, has been defined. Operations that extract propositional information, relative to a particular domain, have also been implemented. For a more complete and detailed list of functions for computational imagery see (Glasgow 1990).

4 Conclusions

This paper treats imagery as a problem solving paradigm in AI and proposes a knowledge representation scheme for computational imagery. Aside from related research in vision, the AI community has given little attention to the topic of imagery. Thus we rely on the relevant theories of cognitive psychology to provide initial guidance for our research. Although psychological plausibility is one of its goals, the scheme was not designed to be a model of human information processing but a tool for practical applications.

The knowledge representation scheme for computational imagery includes three image representations, each appropriate for a different kind of processing: an image is stored in long-term memory as a deep representation that contains relevant descriptive information about the image; the surface representation contains geometric information about the image components; the symbolic array representation is a spatially indexed description of the meaningful parts of the image. A set of primitive functions, which correspond to the fundamental processes involved in mental imagery, have been designed and implemented.

Since the representation scheme is not intended to be a psychological model of mental imagery, we do not suggest that in human working-memory two "mind's eyes" exist that process symbolic and surface representations identical to the ones that we have implemented. What we do claim is that the scheme is informationally equivalent to the representations involved in mental imagery, where two representations are said to be informationally equivalent if all the information in one representation is also inferable from the other and vice versa (Larkin & Simon 1987).

Like Kosslyn's model (Kosslyn 1980), the scheme is based on an array theory of imagery. According to such theories, mental images are array-like surface representations generated through perception, or from long-term, deep representations. Several types of image arrays have previously been suggested. Kosslyn's model uses *2D* arrays for the implementation of the surface representation. 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).

Mental imagery can provide insights that contribute to effective problem solving techniques. A goal of our research is to develop knowledge-based systems that integrate computational imagery with other AI problem solving paradigms. One such system, which is currently under development, is an application to the problem of molecular scene analysis (Glasgow, Fortier and Allen 1991). Other potential applications include vision and tactile perception, medical imaging, motion planning and game playing.

References

Block, N. 1981. *Imagery.* MIT Press.

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.

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

Farah, M.J. 1988. Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review,* 95:307-317.

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.

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. *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.

Jenkins, M.A., Glasgow, J.I., Blevis, E., Chau, R., Hache, E. & Lawson, E. 1988. The Nial AI toolkit. *Proceedings of Avignon '88 Eighth International Workshop on Expert Systems and their Applications.*

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.

Kubovy, M. & Podgorny, P. 1981. Does pattern matching require the normalization of size and orientation? *Perception and Psychophysics,* 30, 24-28.

Larkin, J. & Simon, H.A. 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science,* 10:65-99.

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

Marr, D. & Nishihara, H.K. 1978. Representation and recognition of the spatial organization of three dimensional shapes. *Proceedings of the Royal Society,* B200:269-294.

McDermott, D. V. & Davis, E. 1984. Planning routes through uncertain territory. *Artificial Intelligence,* 22:107-156.

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

Paivio, A. 1975. Perceptual comparisons through the mind's eye. *Memory and Cognition,* 3:635-647.

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.

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.

Samet, H. 1980. Region Representation: Quadtrees from boundary codes. *Comm. of the ACM,* 23: 163-170.