UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Simulating Theories of Mental Imagery

Permalink

https://escholarship.org/uc/item/5mw66768

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 14(0)

Authors

Glasgow, Janice
Conklin, Darrell

Publication Date

1992

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Simulating Theories of Mental Imagery

Janice Glasgow and Darrell Conklin
Department of Computing and Information Science
Queen's University, Kingston
Canada, K7L 3N6
Email: {janice,conklin}@qucis.queensu.ca

Abstract

A knowledge representation scheme for computational imagery has previously been proposed. This scheme incorporates three interrelated representations: a long-term memory descriptive representation and two working memory representations, corresponding to the distinct visual and spatial components of mental imagery. It also includes a set of primitive functions for reconstructing and reasoning with image representations. In this paper we suggest that the representation scheme addresses the controversy involved in the imagery debate by providing the computational tools for specifying, implementing and testing alternative theories of mental imagery. This capability is illustrated by considering the representation and processing issues involved in the mental rotation task.

Introduction

A concept of computational imagery has previously been proposed as a reasoning paradigm in artificial intelligence (Papadias & Glasgow, 1991). The knowledge representation scheme developed for computational imagery can also serve as a tool for specifying, implementing and testing cognitive theories of imagery. Unlike Kosslyn's (1980) computational model, which was designed to support a particular theory, the representations and functions of computational imagery can be used to simulate and analyze alternative, and possibly conflicting, theories of mental imagery.

Computational imagery involves techniques for visual and spatial reasoning, where images are generated or recalled from long-term memory and then manipulated, transformed, scanned, associated with similar forms, increased or reduced in size, distorted, etc. In particular, it is concerned with the reconstruction of image representations to facilitate the retrieval of information that was not explicitly stored in long-term memory. The image representations generated to retrieve this information may correspond to 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 (Glasgow & Papadias, 1992). 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 an image. Each of these representations is constructed, as needed, from a descriptive representation stored in long-term memory. Thus, the 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 using a symbolic array that preserves relevant spatial and topological properties. This array data structure also preserves the hierarchical structure of an image through nested symbolic arrays, which can be used to specify and reason about images at varying levels of the decomposition hierarchy. 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.

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, etc. Because of the relevance of storing and reasoning about such properties of images, the knowledge representation scheme for computational imagery is based on a formal theory of arrays. Similar to set theory, array theory (More, 1979) is concerned with the concepts of nesting and aggregation. It is also concerned with the notion that objects have a location relative to other objects in an array. 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). These functions, along with the primitive functions for computational imagery, provide a general framework for specifying theories of mental imagery.

Computational 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 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).

Pylyshyn (1981), 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. 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 (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. Hinton disputes the picture metaphor for imagery and claims that images are more like generated constructions (Hinton, 1979). In this approach, as in Marr's (1982) *3D* model, complex images can be represented as a hierarchy of parts.

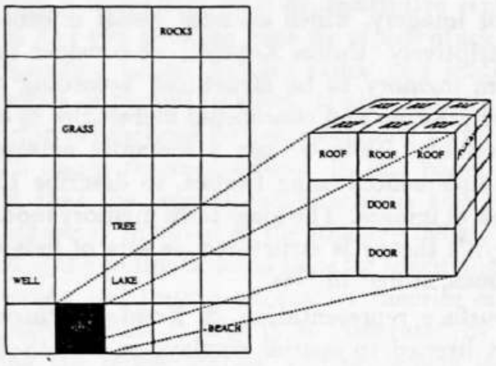
A primary objective of research in cognitive science is to study and explain how the mind functions. One aspect of work in this area is the theory of computability. If a model is computable, then it is usually comprehensible, complete and available for analysis; implementations of theories can be checked for sufficiency and used to simulate new predictive results. In a discussion of the issues of computability of cognitive theories for imagery, Kosslyn (1980) expresses frustration with existing implementation tools and goes on to state that "The ideal would be a precise, explicit language in which to specify the theory and how it maps into the program." Array theory, combined with the primitive functions and representations for computational imagery, provides such a meta-language. Moreover, it allows us to represent an image either visually or spatially, and provides for the implementation and testing of cognitive theories.

In Kosslyn's computational theory for imagery, images have two components: a surface representation (a quasi-pictorial representation that occurs in a visual buffer), and a deep representation for information stored in long-term memory. Similar to Kosslyn, we consider a separate long-term memory model for imagery, which encodes visual information descriptively. Unlike Kosslyn, we consider the long-term memory to be structured according to the decomposition and conceptual hierarchies of an image domain. Thus we use a semantic network model, implemented using frames, to describe the properties of images. The long-term memory model in Kosslyn's theory is structured as sets of lists of propositions, stored in files.

The surface representation in Kosslyn's theory has been likened to spatial displays on a cathode ray tube screen; an image is displayed by selectively filling in cells of a two-dimensional array. Our scheme for representing images in working memory is richer in three fundamental ways. First, we treat images as inherently three-dimensional, although two-dimensional images can be handled as special cases. As pointed out by Pinker (1988), images must be represented and manipulated as patterns in three-dimensions, which can be accessed using either an object-centered or a world-centered coordinate system. Second, we consider two working memory representations, corresponding to the visual and spatial components of mental imagery. Finally, just as the long-term memory stores images hierarchically, the visual and spatial representations use nested arrays to depict varying levels of resolution or abstraction of an image. While the functionality of many of the primitive operations for computational imagery were initially influenced by the processes defined for Kosslyn's theory, their implementation varies greatly because of the nature of the image representations.

An important distinction between our approach to computational imagery and Kosslyn's computational theory is the underlying motivation behind

a) Visual representation



b) Spatial representation

Figure 1: Representations for computational imagery

the two pieces of work. Kosslyn's model was initially developed to simulate and test a particular theory for mental imagery. Although in its initial development the representations and processes of computational imagery were motivated by efficiency and expressiveness concerns, it was also designed to provide the predictive and explanatory power necessary to model alternative theories of cognition.

As an illustration of the visual and spatial representations for computational imagery, consider the island map used by Kosslyn to investigate the processes involved in mental image scanning. Figure 1 presents an occupancy array visual depiction of such a map, as well as a symbolic array spatial representation that preserves the properties of closeness (expressed as adjacency) and relative location of the important features of the island. It can also be used to depict and reason about the hierarchical structure of images (see subimage for hut) using nested data structures. Consider addressing such questions as those involved in mental image scanning and focusing of attention. Although both representations provide access to information about relative location and could be used to simulate scanning processes, they would predict different results in terms of time complexity.

The hypothesis of distinct visual and spatial image components of mental imagery is suggested by results of studies of visually impaired patients (e.g. Kosslyn 1987). It is further supported by the conflicting results of studies dealing with whether size or relative distance is preserved in the representation of a mental image. This controversy could be attributed to the hypothesis of multiple image representations. Thus, as well as testing theories involving individual representations, computational imagery can be used to analyze theories that involve multiple representations.

It is worth noting here that the spatial representation for computational imagery is not just an interpreted low resolution version, or approximation, of the visual representation of an image. The symbolic array may discard, not just approximate, irrelevant visual information. It may also abstract away details by using nested arrays to depict image components at varying levels of the structural hierarchy.

Another approach to visual reasoning was presented by Funt (1980), who represented the state of the world as a diagram, and actions in the world as corresponding actions in the diagram. Similar to Kosslyn, Funt uses two-dimensional arrays to denote his visual images. A more recent model 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). While this model allows a hierarchy of descriptions, it is not spatially organized.

The subjective nature of mental imagery has made it a difficult topic to study experimentally; many qualities of an image are not directly observable and may differ from one person to another. In fact, it has been argued by some researchers that it is impossible to resolve the imagery debate experimentally (Anderson, 1978). The difficulties involved in psychological studies emphasize the importance of computer models for mental imagery. While computational imagery does not imply a particular model, it does provide the tools for formally analyzing a variety of theories, and thus may contribute to resolving the imagery debate.

Theories of Rotation

To illustrate the use of computational imagery for studying cognitive theories of imagery, we consider the problem of mental rotation. Although empirical observations suggest that rotation involves an object representation being moved through intermediate orientations (Shepard & Cooper, 1982), an as yet unresolved issue is the actual nature of the representation used. One possibility is a visual depiction of the object which preserves detailed three-dimensional shape information. An alternative approach is one in which the object is represented as vectors corresponding to the major skeletal axes of the object (Just & Carpenter, 1985). This type of representation can be considered as spatial in nature; it preserves connectivity of parts but discards detailed surface information about the image.

A visual representation of an object can be used to test theories of mental rotation where an object is incrementally shifted through some medium. The medium, such as occupancy arrays (Kosslyn, 1980) or concentric rings in a simulated retina structure (Funt, 1983), supports visual inferences and transformations. An object, or a portion of an object, is rotated by shifting its position in the medium. As Funt illustrates, this can be done in parallel using message passing between adjacent cells in a retinal diagram. In Kosslyn's rectangular model, objects will be deformed as they are incrementally rotated, and the granularity of the medium is one of the determining factors of rotation time. This approach can similarly be implemented using a three-dimensional occupancy array in our representation scheme.

The spatial representation of an object preserves important relations among parts, and is pseudo-visual in that it does not necessarily preserve relative distance or metric information. A theory of rotation using symbolic arrays would have an encoding followed by a comparison process. First, an object is rotated into a Cartesian axis-aligned orientation. It is then scanned to extract the skeletal axes, which are labeled and encoded into a symbolic array. Figure 2 shows the partial encoding grammar for simple two-dimensional orthogonal figures. On the left is a visual depiction of such a figure, in the middle its labeled skeletal axes, and on the right the symbolic spatial array. For brevity we do not show the encoding for all four orientations of each stick figure. Encoding visual objects using this grammar preserves such important relations as connectivity and ternary symmetry. Figure 3 shows the encodings for more complex visual objects. In this figure, (A) and (B) denote the same image at varying orientations, while (C) is a mirror image of the first two figures.

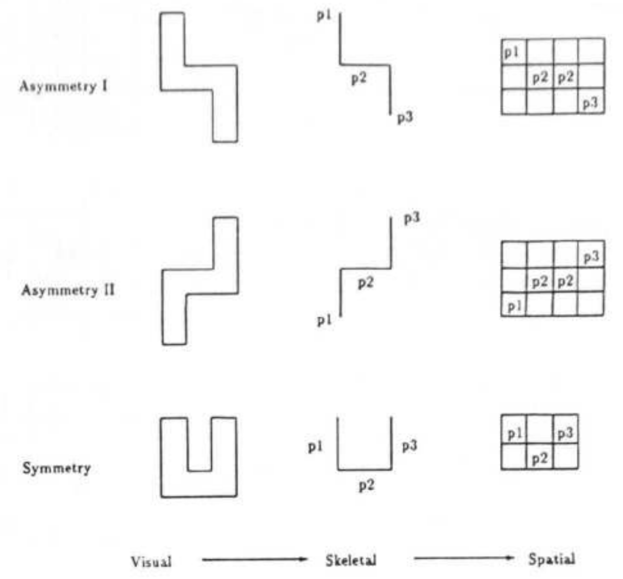


Figure 2: Encoding connected regions of a visual object in a symbolic array.

Two symbolic arrays are considered equivalent under a particular relation if there is a relation homomorphism between them. Informally, this is a function, f, which maps the parts of one symbolic array to the parts of another such that if parts are related in one array, the function maps them to parts which are similarly related. For example, two symbolic arrays $A1$ and $A2$ are equivalent under the relation of connectivity if and only if there exists a one-to-one mapping f between the components of $A1$ and $A2$ such that for all a, b in $A1$, $connected(a,b) \leftrightarrow connected(f(a), f(b))$. Spatial relations are determined simply by computationally inspecting the symbolic array representation. Consider the two symbolic arrays in (A) and (B) of Figure 3. There exists a mapping $f = \{(p1, q5), (p2, q4), (p3, q3), (p4, q2), (p5, q1)\}$ which preserves the relation of connectivity. This mapping also preserves the ternary symmetry relations specified in Figure 2. Although the array in (C) is equivalent under the connectivity relation to the other two, there does not exist a mapping in which the symmetry relations are preserved.

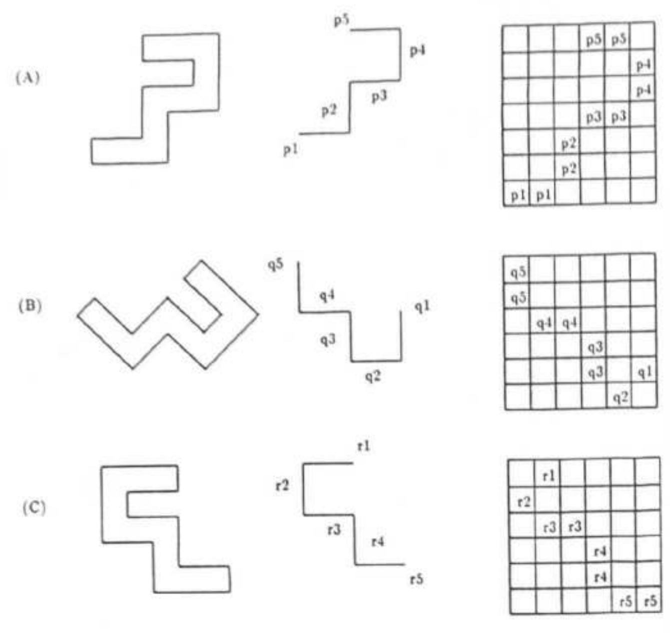


Figure 3: Spatial encodings for visual objects

For this simple class of orthogonal figures, we rephrase the problem of rotation as one of determining whether two images are equivalent under the relations of symmetry and connectivity. The interesting point is that testing two spatial representations for equivalence under these relations involves only identifying a terminal segment for each object, tracing the connectivity of the object in the symbolic array, while checking that symmetry is maintained. It involves no explicit rotation or transformation of the arrays. Note that the symbolic arrays for (A) and (B) in Figure 3 satisfy the conditions for rotation, whereas the array in (C) is not spatially equivalent to the other two arrays.

This spatial rotation method might explain the comparison of *orientation-free* descriptions reported by Just and Carpenter (1985), who hypothesized, based on experimental results, that rotation can sometimes be performed without explicit object manipulation. In their experiments on the comparison of cubes, a subject using an orientation free description strategy exhibited rotation times more or less constant and independent of the amount of rotation involved.

In this section we have presented one example illustrating the use of the spatial representation for computational imagery to explain a mental rotation result. Other issues we could use our knowledge representation scheme to address include whether objects are rotated in whole or in parts, parallel versus sequential process models, axis or task-defined axes for visual objects (Just and Carpenter, 1985), and the use of hierarchical encoding schemes (Anderson, 1983). Similarly, we could examine alternative theories for tasks involving mental scanning, creative imagery and the relationship between imagery and vision.

Discussion

Previous research in imagery has focussed on two issues: the nature of the image representation and the function of imagery in tasks such as rotation, scanning and language processing. The proposed knowledge representation scheme for computational imagery allows us to address both of these interrelated issues; image representations are data structures which can be considered in conjunction with, or distinctly from, the functions that operate on them. The multiple representations of the scheme provides a computational framework for considering images as propositions stored in long-term memory, or as working memory visual or spatial depictions. Using the primitive functions for computational imagery, we can study the relationship between mind and computer by implementing and testing algorithms that simulate theories of mental imagery.

The knowledge representation scheme for computational imagery also provides a basis for implementing programs that involve reconstructing and reasoning with image representations (Glasgow, 1990). In particular, it can be used to develop knowledge-based systems for problems that involve mental imagery when solved by humans. One such system, which is currently under development, is an application to the problem of molecular scene analysis (Glasgow, Fortier & Allen, 1991). This knowledge-based system combines tools form the areas of protein crystallography and molecular-database analysis, through a computational imagery framework. The spatial representation for computational imagery has also been applied to the problem of understanding the use of diagrams and visual analogies in the development of Dalton's atomic theory (Thagard & Hardy, 1992).

Computational imagery is currently being considered in the domain of machine learning (Conklin & Glasgow, 1992). In this research, the spatial representation for imagery is used to develop a theory of subsumption and similarity, which provides a formal model for structure-based classification in the long-term memory model of images.

In conclusion, the underlying mathematics for computational imagery satisfies Kosslyn's (1980) ideal by providing a precise and explicit language for specifying theories of mental imagery. Visual and spatial representations are implemented as arrays and manipulated using the primitive functions of computational imagery, which themselves are expressed as array theory operations. Finally, the primitives of array theory, computational imagery and particular theories of mental imagery can all be directly mapped into programs which run without the "kluges" and "ad hoc manipulations" faced by Kosslyn in the development of his computational theory.

References

Anderson, J.R. 1978. Arguments concerning representations for mental imagery. *Psychological Review,* 85, 249-277.

Block, N., editor 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, 670-677.

Conklin, D. & Glasgow, J.I. 1992. Spatial analogy and subsumption. In Sleeman & Edwards (Eds.), *Machine Learning: Proceedings of the Ninth International Conference (ML92).* Morgan Kaufmann.

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.

Funt, B.V. 1983. A parallel-process model of mental rotation. *Cognitive Science,* 7, 67-93.

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, Forthcoming.

Glasgow, J.I. & Papadias, D. 1992. Computational Imagery, *Cognitive Science,* Forthcoming.

Hinton, G. 1979. Some demonstrations of the effects of structural descriptions in mental imagery, *Cognitive Science,* 3, 231-250.

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

Just, M.A. & Carpenter, P.A. 1985. Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review,* 92, 137-172.

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.

Marr, D. 1982. *Vision: a computational investigation in the human representation of visual information.* San Francisco: Freeman.

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. 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. 1981. The imagery debate: analogue media versus tacit knowledge. *Psychological Review,* 88, 16-45.

Shepard, R.N. & Cooper, L.A. 1982. *Mental images and their transformations.* MIT Press.

Thagard, P. & Hardy, S. 1992. Visual thinking in the development of Dalton's atomic theory, *Proceedings of AI '92.*