Strong spatial cognition

Christian Freksa
University of Bremen, Germany

Motivation

The ability to solve spatial tasks is crucial for everyday life and thus of great importance for cognitive agents. A common approach to modeling this ability in artificial intelligence has been to represent spatial configurations and spatial tasks in form of *knowledge about* space and time. Augmented by appropriate algorithms such representations allow the computation of knowledge-based solutions to spatial problems. In comparison, natural embodied and situated cognitive agents often solve spatial tasks without detailed knowledge about underlying geometric and mechanical laws and relationships; they can directly relate actions and their effects due to spatio-temporal affordances inherent in their bodies and their environments. Against this background, we argue that spatial and temporal structures *in the body and the environment* can substantially support (or even replace) reasoning effort in computational processes. While the principle underlying this approach is well known—for example, it is applied in descriptive geometry for geometric problem solving—it has not been investigated as a paradigm of cognitive processing. The relevance of this principle may not only be to overcome the need for detailed knowledge that is required for a knowledge-based approach; it is also in understanding the efficiency of natural problem solving approaches.

Architecture of cognitive systems

Cognitive agents such as humans, animals, and autonomous robots comprise brains (resp computers) connected to sensors and actuators. These are arranged in their (species-specific) bodies to interact with their (species-typical) environments. All of these components need to be well tuned to one another to function in a fully effective manner. For this reason, it is appropriate to view the entire aggregate (cognitive agent including body and environment) as a 'full cognitive system' (Fig. 1).

Our work aims at investigating the distribution, coordination, and execution of tasks among the system components of embodied and situated spatial cognitive agents. From a classical information processing/AI point of view, the relevant components outside the brain or computer would be formalized in some knowledge representation language or associated pattern in order to allow the computer to perform formal reasoning or other computational processing on this representation. In effect, physical, topological, and geometric relations are transformed into abstract *information* about these relations and the tasks are then performed entirely on the information processing level, where true physical, topological, and geometric relations no longer persist.

This classical information-processing oriented division between brain/computer on one hand and perception, action, body, and environment on the other hand is only one way of distributing the activities involved in cognitive processing [Wintermute and Laird, 2008]. Alternative ways would be (1) to maintain some of the spatial relations in their original form or (2) to use only 'mild abstraction' for their representation. Maintaining relations in their original form corresponds to what Norman [1980] named *knowledge in the world.* Use of knowledge in the world requires perception of the world to solve a problem. The best-known example of mild abstraction is geographic paper maps; here certain spatial relations can be represented by identical spatial relations (e.g. orientation relations); others could be transformed (e.g. absolute distances could be scaled). As a result, physical operations such as perception, route-following with a finger, and manipulation may remain enabled similarly as in the original domain. Again, perception is required to use these mildly abstracted representations—but the perception task can be easier than the same task under real-world conditions, for example due to the modified scale.

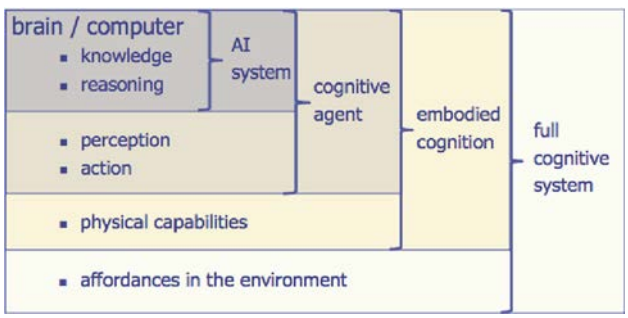

Fig. 1 Structure of a full cognitive system

A main research hypothesis for studying physical operations and processes in spatial and temporal form in comparison to formal or computational structures is that spatial and temporal structures *in the body and the environment* can substantially support reasoning effort in computational processes. One major observation we can make when comparing the use of such different forms of representation (formal, mild abstraction, original) is that the processing structures of problem solving processes differ [Marr 1982]. Different processing structures facilitate different ease of processing [Sloman 1985].

Our hypothesis can be plainly formulated as:

manipulation + perception simplify computation

While the principle underlying this hypothesis is well known—for example, it is applied in descriptive geometry for geometric problem solving—it has not been investigated as a principle of cognitive processing.

Reasoning about the world can be considered the most advanced level of cognitive ability; this ability requires a comprehensive understanding of the mechanisms responsible for the behavior of bodies and environments. But many natural cognitive agents (including adults, children, and animals) lack a detailed understanding of their environments and still are able to interact with them rather intelligently. For example, they may be able to open and close doors in a goal-directed fashion without understanding the mechanisms of the doors or locks on a functional level. This suggests that knowledge-based reasoning may not be the only way to implementing problem solving in cognitive systems.

In fact, alternative models of perceiving and moving goal-oriented autonomous systems have been proposed in biocybernetics and AI research to model aspects of cognitive agents [e.g. Braitenberg 1984; Brooks 1991; Pfeifer and Scheier, 2001]. These models physically implement perceptual and cognitive mechanisms rather than describing them formally and coding them in software. Such systems are capable of intelligently dealing with their environments without encoding knowledge about the mechanisms behind the actions.

The background of the present work has been discussed in detail in [Freksa 2013; Freksa and Schultheis, in press].

Approach

With our present work, we go an important step beyond previous embodied cognition approaches to spatial problem solving. We introduce a paradigm shift which not only aims at preserving spatial structure, but also will make use of identity preservation; in other words, we will represent spatial objects and configurations by themselves or by *physical spatial* models of themselves, rather than by abstract representations. This has a number of advantages: we can avoid loss of information due to early representational commitments: we do not have to decide prematurely which aspects of the world to represent and which aspects to abstract from. This can be decided partly during the problem solving procedure. At this stage, additional contextual information may become available that can guide the choice of the specific representation to be used.

Perhaps more importantly, objects and configurations frequently are aggregated in a natural and meaningful way; for example, a chair may consist of a seat, several legs, and a back; if I move one component of a chair, I automatically (and simultaneously!) move the other components and the entire chair, and vice versa. This property is not intrinsically given in abstract representations of physical objects; but it may be a very useful property from a cognitive point of view, as no computational processing cycles are required for simulating the physical effects or for reasoning about them. Thus, manipulability of physical structures may become an important feature of cognitive processing, and not merely a property of physical objects.

Similarly, we aim at dealing with perception dynamically, for example allowing for "on-the-fly" creation of suitable spatial reference frames: by making direct use of spatial configurations, we can avoid deciding a *priori* for a specific spatial reference system in which to perceive a configuration. As we know from problem solving in geometry and from spatial cognition, certain reference frames may allow a spatial problem to collapse in dimensionality and difficulty. For example, determining the shortest route between two points on a map boils down to a 1-dimensional problem [Dewdney 1988]. However, it may be difficult or impossible to algorithmically determine a reference frame that reduces the task given on a 2- or 3-dimensional map to a 1-dimensional problem. A spatial reconfiguration approach that makes use of the physical affordance 'shortcut', easily reduces the problem from 3D or 2D to 1D. In other cases, it may be easier to identify suitable spatial perspectives empirically *in the field* than analytically by computation. Therefore we may be better off by allowing certain operations to be carried out situation-based in the physical spatial configuration as part of the overall problem solving process.

In other words, our project investigates an alternative architecture of artificial cognitive systems that may be more closely based on role models of natural cognitive systems than our purely knowledge-based AI approaches to cognitive processing. We focus on solving spatial and spatio-temporal tasks, i.e. tasks having physical aspects that are directly accessible by perception and can be manipulated by physical action. This will permit 'outsourcing' some of the 'intelligence' for problem solving into spatial configurations.

Our approach is to first isolate and simplify the specific spatial problem to be solved, for example by identifying an appropriate task-specific spatial reference system, by removing task-irrelevant entities from the spatial configuration, or by reconstructing the essence of the spatial configuration by minimal abstraction. In general, it may be difficult to prescribe the precise steps to preprocess the task; for the special case of spatial tasks it will be possible to provide rules or heuristics for useful preprocessing steps; these can serve as meta-knowledge necessary to control actions on the physical level. After successful preprocessing, it may be possible in some cases to 'read' an answer to the problem through perception directly off the resulting configuration; in other cases the resulting spatial configuration may be a more suitable starting point for a knowledge-based approach to solving the problem.

Discussion

The main hypothesis of our approach is that the 'intelligence' of cognitive systems is located not only in specific abstract problem-solving approaches, but also—and perhaps more importantly—in the capability of recognizing characteristic problem structures and of selecting particularly suitable problem-solving approaches for given tasks. Formal representations may not facilitate the recognition of such structures, due to a bias inherent in the abstraction. This is, where *mild abstraction* can help: mild abstraction may abstract only from few aspects while preserving important structural properties.

The insight that spatial relations and physical operations are strongly connected to cognitive processing may lead to a different division of labor between the perceptual, the representational, the computational, and the locomotive parts of cognitive interaction than the one we currently pursue in AI systems: rather than putting all the 'intelligence' of the system into the computer, the proposed approach aims at putting more intelligence into the interactions between components and structures of the full cognitive system. More specifically, we aim at exploiting intrinsic structures of space and time to simplify the tasks to be solved.

We hypothesize that this flexible assignment of physical and computational resources for cognitive problem solving may be closer to natural cognitive systems than the almost exclusively computational approach; for example, when we as cognitive agents search for certain objects in our environment, we have at least two different strategies at our disposal: we can represent the object in our mind and try to imagine and mentally reconstruct where it could or should be—this would correspond to the classical AI approach; or we can visually search for the object in our physical environment. Which approach is better (or more promising) depends on a variety of factors including memory and physical effort; frequently a clever combination of both approaches may be best.

Although the general principle outlined may apply to a variety of domains, we will constrain our work in the proposed project to the spatio-temporal domain. This is the domain we understand best in terms of computational structures; it has the advantage that we have well-established and universally accepted reference systems to describe and compute spatial and temporal relations.

Our research aims at identifying a bag of cognitive principles and ways of combining them to obtain cognitive performance in spatio-temporal domains. We bring together three different perspectives, in this project: (1) the cognitive systems perspective which addresses cognitive architecture and trade-offs between explicit and implicit representations; (2) the formal perspective which characterizes and analyzes the resulting structures and operations; and (3) the implementation perspective which constructs and explores varieties of cognitive system configurations. In the long-term, we see potential technical applications of physically supported cognitive configurations for example in the development of future *intelligent materials* (e.g. 'smart skin' where distributed spatio-temporal computation is required but needs to be minimized with respect to computation cycles and energy consumption).

Naturally, the proposed approach will not be as broadly applicable as some of the approaches we pursue in classical AI. But it might discover broadly applicable cognitive engineering principles, which will help the design of tomorrow's intelligent agents. Our philosophy is to understand and exploit pertinent features of space and time as modality-specific properties of cognitive systems that enable powerful specialized approaches in the specific domain of space and time. However, space and time are most basic for perception and action and ubiquitous in cognitive processing; therefore we believe that understanding and use of their specific structures may be particularly beneficial.

In analogy to the notion of 'strong AI' (implementing intelligence rather than simulating it [Searle 1980]) we call this approach 'strong spatial cognition', as we employ real space rather than simulating its structure.

Acknowledgments

I acknowledge discussions with Holger Schultheis, Ana-Maria Olteteanu, and the R1-[ImageSpace] project team of the SFB/TR 8 Spatial Cognition. This work was generously supported by the German Research Foundation (DFG).

References

Braitenberg V (1984) Vehicles: experiments in synthetic psychology. MIT Press, Cambridge

Brooks RA (1991) Intelligence without representation, Artif Intell 47:139–159

Dewdney AK (1988) The armchair universe. W.H. Freeman & Company, San Francisco
Freksa C (2013) Spatial computing—how spatial structures replace computational effort. In: Raubal M, Mark D, Frank A (eds) Cognitive and linguistic aspects of geographic space. Springer, Heidelberg
Freksa C, Schultheis H (in press) Three ways of using space. In: Montello DR, Grossner KE, Janelle DG (eds) Space in mind: concepts for spatial education. MIT Press, Cambridge
Marr D (1982) Vision. MIT Press, Cambridge
Norman DA (1980) The psychology of everyday things. Basic Books, Inc, New York
Pfeifer R, Scheier C (2001) Understanding intelligence. MIT Press, Cambridge
Searle J (1980) Minds, brains and programs. Behav Brain Sci 3(3): 417–457
Sloman A (1985) Why we need many knowledge representation formalisms. In Bramer M (ed) Research and development in expert systems. Cambridge University Press, New York, pp 163–183
Wintermute S, Laird JE (2008) Bimodal spatial reasoning with continuous motion. In: Proceedings of AAAI, pp 1331–1337

Inferring 3D shape from texture: a biologically inspired model architecture

Olman Gomez, Heiko Neumann
Inst. of Neural Information Processing, Ulm University, Germany

Abstract
A biologically inspired model architecture for inferring 3D shape from textures is proposed. The model is hierarchically organized into modules roughly corresponding to visual cortical areas in the ventral stream. Initial orientation selective filtering decomposes the input into low-level orientation and spatial frequency representations. Grouping of spatially anisotropic orientation responses builds sketch-like representations of surface shape. Gradients in orientation fields and subsequent integration infers local surface geometry and globally consistent 3D depth.

Keywords
3D Shape, Texture, Gradient, Neural Surface Representation

Introduction
The representation of depth structure can be computed from various visual cues such as binocular disparity, kinetic motion and texture gradients. Based on findings from experimental investigations (Liu et al. (2004); Tsutsui et al. (2002)) we suggest that depth of textured surfaces is inferred from monocular images by a series of processing stages along the ventral stream in visual cortex. Each of these stages is related to individual cortical areas or a strongly clustered group of areas (Markov et al. 2013). Based on previous works that develop generic computational mechanisms of visual cortical network processing (Thielscher and Neumann (2003); Weidenbacher et al. (2006)) we propose a model that transforms initial texture gradient patterns into representations of intrinsic structure of curved surfaces (lines of minimal curvature, local self- occlusions) and 3D depth (Li and Zaidi (2000); Todd (2004)).

Previous work
Visual texture can assume different component structure which suffers from compression along the direction of surface slant when the object appearance curves away from the viewer's sight. Texture gradients provide a potent cue to local relative depth (Gibson, 1950). Several studies have investigated how size, orientation or density of texture elements convey texture gradient information (Todd and Akerstrom, 1987). Evidence suggests that patterns of changing energy convey the basic information to infer shape from texture that need to be integrated along characteristic intrinsic surface lines (Li and Zaidi, 2000). Previous computational models try to estimate surface orientation from distortions of the apparent optical texture in the image. The approaches can be subdivided according to their task specificity and the computational strategies involved. Geometric approaches are suggested to reconstruct the structure of the metric surface geometry (e.g., Aloimonos and Swain (1985); Bajcsy and Lieberman (1976); Super and Bovik (1995)). Neural models, on the other hand, infer the relative or even ordinal structure from initial spatial frequency selective filtering, subsequent grouping of the resulting output responses and a depth mapping step (Grossberg et al. 2007; Sakai and Finkel, 1997). The LIGHTSHAFT model of Grossberg et al. (2007) utilizes scale-selective initial orientation filtering and subsequent long-range grouping. Relative depth in this model is inferred by depth-to-scale mapping associating coarse-to-fine filter scales to depth using orientation sensitive grouping cells which define scale- sensitive spatial compartments to fill-in qualitative depth. Grouping mechanisms can be utilized to generate a raw surface sketch to establish lines of minimal surface curvature as a ridge-based qualitative geometry representation (Weidenbacher et al. 2006). Texture gradients can be integrated to derive local maps of relative surface orientation (as suggested in Li and Zaidi (2000); Sakai and Finkel (1997)). Such responses may be integrated to generate globally consistent relative depth maps from such local gradient responses (Liu et al. 2004).

The above mentioned models are limited to simple objects most dealing only with regular textures and do not give an explanation as to how the visual system mechanistically produces a multiple depth order representation of complex objects.

Model description
Our model architecture consists of a multi-stage network of interacting areas that are coupled bidirectionally (extension of (Weidenbacher et al. 2006); Fig. 1). The architecture is composed of four functional building blocks or modules, each one consists of three stages corresponding to the compartment structure of cortical areas: feedforward input is initially filtered by a mechanism specific to the model area, then resulting activity is modulated by multiplicative feedback signals to enhance their gain, and finally a normalization via surround competition utilizes a pool of cells in the space-feature domain.

The different stages can be formally denoted by the following steady-state equations (with the filter output modulated by feedback and inhibition by activities from a pool of cells (Eq. 1) and the inhibitory pool integration (Eq. 2)):

$$r^{I}_{i,feat} = \frac{\beta \cdot f(F(r^{0})) \cdot \left(1 + net^{I,FB}_{i,feat}\right) - \xi \cdot q^{I,in}_{i,feat} + \eta}{\alpha + \gamma \cdot f(F(r^{0})) \cdot \left(1 + net^{I,FB}_{i,feat}\right) + q^{I,in}_{i,feat}} \tag{1}$$

$$q^{I,in}_{i,feat} = \delta \cdot \left(\sum_{feat} r^{I}_{i,feat} + \varepsilon \cdot \sum_{j} \max_{feat} \left(r^{I}_{i,feat} \right) \cdot \Lambda^{pool}_{ij} \right) \tag{2}$$

where the feedback signal is defined by $net^{I,FB}_{i,feat} = [\lambda_{FB} - r^{II}_{i,feat}] + \sum_{z \in \{feat,loc\}} r^{II}_{z}$. Here r^{I}, r^{II} denote output activation of the generic modules (I, II: two subsequent modules in the hierarchy). The different three-stage modules roughly correspond to different cortical areas with different feature dimensions represented neurally (compare Fig. 1): Cortical area V1 computes orientation selective responses using a spatial frequency decomposition of the input; area V2 accomplishes orientation sensitive grouping of initial items into boundaries in different frequency channels to generate representations of surface curvature properties. Different sub-populations of cells in V4/IT are proposed to detect different surface features from distributed responses: One is used to extract discontinuities in the orientation fields (indicative for self-occlusions), another extracts and analyzes anisotropies in the orientation fields of grouping responses to