Sketch Learning by Analogy

Ulf KRUMNACK [a], Angela SCHWERING [b], Kai-Uwe KÜHNBERGER [a], Helmar GUST [a], Ahmed ABDEL-FATTAH [a,1], Tarek BESOLD [a], Martin SCHMIDT [a], and Stefan SCHNEIDER [a]

[a] *Institute of Cognitive Science, University of Osnabrück, Germany*

[b] *Institute for Geoinformatics, University of Münster, Germany*

Abstract. Sketches are shapes that represent objects, scenes, or ideas by depicting relevant parts and their spatial arrangements. While humans are quite efficient in understanding and using sketch drawings, those are largely inaccessible to computers. We argue that this is due to a specific shape based representation by humans and hence the use of cognitively inspired representation and reasoning techniques could lead to more proficient sketch processing. We also propose a three-level architecture for sketch learning and recognition that builds on concepts from cognitive science, especially from analogy research, to map and generalize sketches.

Keywords. Sketch, Shape, Learning, Analogy

1. Introduction

Sketches can be considered as an intermediate level of abstraction between raw sub-symbolic streams of sensory input on the one side and icons on the other. In contrast to a drawing, a sketch only captures the conceptually relevant parts of the displayed object or situation as well as the spatial relations between these parts, making their treatment substantially different from classical image processing. The pertinence of sketches for future information technology applications and services can hardly be overestimated. Especially the spread of tablet computers and devices equipped with touch screens paves the way for new human computer interfaces, in which sketches can play an essential role. Future applications can be search services for large knowledge bases utilizing input sketches, support services in software systems for shortening the path through complex menus, automatic sketch generation for manuals and assembly instructions, a bridging approach between computer vision and conceptual reasoning, or creative usage of sketches in e-learning contexts.

In this paper, we present ideas on modeling the human ability to operate with sketches. We focus on a competence model for recognition, classification, memorization and retrieval of sketches guided by cognitive principles. In a first step, the envisaged system acquires basic knowledge on how to sketch a given

[1] Authors are listed in alphabetical order.

object. The essential and optional components as well as their spatial arrangement are learned by comparing different sketches of the same type of object provided to the system as training data. In the next step, after elementary types have been learned in this bootstrapping process, the system will generate more abstract categories by cross-type comparison, establishing a hierarchical index of sketch schemata and shapes. This index will then support the recognition capacity: new sketches will be compared to the abstract descriptions in the sketch database to find structurally matching sketches in memory. We argue in favor of a symbolic approach because the structure of a sketch can be captured explicitly in such a representation, and changes in the conceptualization can be performed by automatic inference techniques.

The paper is structured as follows. We start with discussing requirements for a representation language for sketches in section 2. The description of the proposed system is given in section 3, which constitutes the main part of this paper. We then provide links to related work in section 4, before concluding with some remarks and future work in section 5.

2. Sketch Representation and Re-representation

Sketches are assumed to be given as a collection of dots and lines, possibly annotated with an order of drawing. Multiple relational representations can thus be constructed based on psychological principles, which take into account that human cognition of spatial environments is qualitative in nature. Humans do not perceive absolute locations and quantitative relations between spatial objects, but rather relative locations and qualitative relations [1,2,3,4]. By observing a geometric figure, the unstructured information is transformed into a structured representation of coherent shapes and patterns [5,6]. Perception tends to follow a set of Gestalt principles: stimuli are experienced as a possibly good Gestalt, i.e. as regular, simplistic, ordered, and symmetrical as possible. Gestalt psychology argues that human perception is holistic: instead of collecting every single element of a spatial object and afterwards composing all parts into one integrated picture, people experience things as an integral, meaningful whole. The whole contains an internal structure described by relationships among the individual elements.

We argue that qualitative spatial relations play a major role during sketch recognition and hence sketches should be described on a qualitative level by a symbolic language. The spatial representation language has to meet two major requirements: it must describe all elements of a spatial object with respect to the aspects relevant in human perception, and it must also describe the spatial characteristics that are important in recognizing spatial objects. To reflect human perception, the language must comprise significant perceptual vocabulary to specify visual structures. The geometry in a sketch, i.e. of its elements and their spatial relations, has to be represented in a way that allows for cognitively plausible reasoning. The language can be based on psychological theories for perception and pattern recognition, such as Gestalt Theory [7,8,5,6], Marr's theory of vision [9] and Biedermann's Geons [10], and on research specifically directed towards the sketch mapping task such as the CogSketch [11] approach.

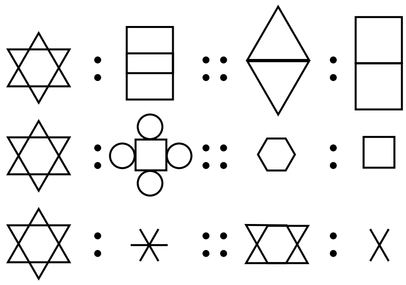

Figure 1. Visual ambiguity exemplified by different representations of a sketch relative to a given context in a proportional analogy, according to [12].

The potential ambiguity of sketches, e.g. caused by different groupings of elements or different interpretations, is an essential point to be considered. Indurkhya [12] has demonstrated the effects of visual ambiguity in proportional geometric analogies and has argued for a mechanism that can change representations. The importance of re-representation is exemplified in Figure 1, where structural commonalities between representations can be detected only if suitable representations for the geometric figure are available. The Star of David in the top row of Figure 1 should be represented as two overlapping triangles, whereas the one in the middle row should be represented as six triangles plus a central hexagon, and that in the bottom row should be represented as three overlapping rhombuses. Re-representation in this case means changing from one of these representations to another one which suits better to the given problem.

Re-representation, in the domain of sketches, means spatial re-organization and re-structuring of the elements within a spatial object, and can be formalized as a deduction task: from a given description of a sketch an alternative description has to be derived, that represents the same visual scene. It therefore requires spatial reasoning capabilities and existing qualitative spatial reasoners can be used to support this task (such as the SparQ toolbox [13] or General Qualitative Reasoner (GQR) [14]). Furthermore, to reflect human strategies of re-representation, appropriate heuristics are needed to guide the re-representation process.

3. A System for Analogy-Based Sketch Learning

Human learning is not a one-step action but a continuous, incremental process of acquiring new and revising old knowledge, where knowledge is learned at different levels of abstraction. Such observations about human learning motivate us to develop a three-level architecture for learning perceptual categories based on sketches. Perceptual categories in this context refer to structured representations of graphical elements that are common to a class of sketch drawings, represented as structured descriptions with respect to relevant topological, directional, and geometrical properties. The two main mechanisms for learning are learning via transfer and learning by abstraction. The former refers to the transfer of facts from the source to the target domain, while the latter denotes the generalization process that is essential to derive abstract concept definitions. Existing classical learning approaches usually require large sets of data samples to create generalizations, though humans can already generalize over a small set of samples.

Our proposed system applies analogical comparison to discover structural commonalities and combines them with inductive refinement to extract the essential characteristics defining a perceptual category. Analogy-making, as a non-standard reasoning technique, is combined with classical deductive and inductive reasoning to compare different sketch drawings for commonalities and generalize a common underlying perceptual category. For all tasks involving comparison of sketches, analogical mapping is used to align two stimuli based on structural similarities. Such a mapping is essentially shape based, i.e. it is performed on visual descriptions only, and does not rely on functional, intensional, or usage-based information. There are two central requirements that need to be realized. The system needs, first, to be able to incrementally add newly learned categories, and, secondly, to be adaptive in the sense that a computed generalization is modifiable if new stimuli require a relaxation of the imposed constraints. Knowledge learned from training examples can be used to recognize and classify new sketches.

The model presented in this section is inspired by [15], where first ideas for an incremental learning theory were proposed. In that paper, we used a multi-layered model based on analogies to explain how abstract physical principles such as the law of energy conservation and the concept of an equilibrium of forces can be learned. These ideas are revived here and applied to the domain of sketches yielding a three-level architecture. The first level refers to the computation of analogical generalizations between a pair of sketches (section 3.1). The second level is the inductive refinement of the computed generalizations based on a re-representation process that adapts representations to make it compatible to further sketches (section 3.2). The third level focuses on learning through a revision process when comparing abstract generalizations to new domains (section 3.3). Finally, we discuss how the aquired knowledge can be used for sketch recognition (section 3.4).

3.1. Level 1: Analogical Generalization

At the lowest level, two sketches are taken as input, and an analogy between them is computed based on structural commonalities (cf. Figure 2). The relational structure of the description of the sketches is thus crucial. The analogical mapping may be partial, i.e. it allows parts of one sketch that have no counter-parts in

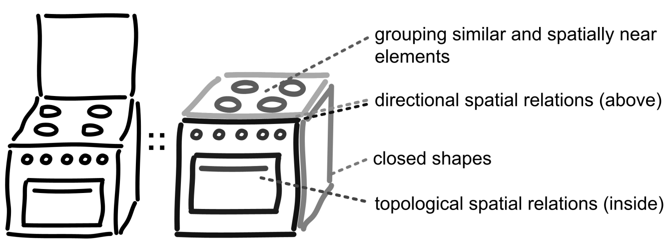

Figure 2. A flat description of a sketch is mapped to a structural representation

the other sketch. The mapping will give rise to a generalization, i.e. an abstract description of the common parts of both sketches.

Heuristic-driven theory projection (HDTP) is a logic-based framework for analogy making, presented in [16], where domains are described by logical theories and are represented by a finite set of axioms. An analogy is established by mapping axioms of two domains, based on a generalization computed via anti-unification (cf. [17]). HDTP allows re-representation of input domains: If the axiomatizations provided for the domains do not exhibit sufficient common structure to establish a good analogy, formulas from the domain theory, which can be derived from the axioms by logical deduction, are considered for mapping (cf. [18]).

The framework uses a set of heuristics to compute an analogical mapping that can be adapted to fit the special needs in the sketch domain. Essential complexity measures and heuristics are applied on different levels to guide the alignment process and to evaluate possible mappings in the sketch mapping scenario. Heuristics are used to (1) determine the order in which axioms are selected and included in the mapping process: psychologically motivated (and syntactic) heuristics can proof useful, where perceptually significant elements in human perception are likely to influence the analogy-making process more than non-significant elements (axioms should be selected therefore in the order of perceptual significance); (2) guide the re-representation: heuristics should reflect human strategies of re-representation, and the spatial language, particularly the re-representation rules, influences the development of the heuristics; and (3) determine when an analogy contains sufficient analogous structures such that a new sketch stimulus can be classified as a certain object. The approach has to bridge the gap between largest possible mappings – the more analogical structures are identified, the better the analogy – and differences in the sketches that should not be part of the analogy.

3.2. Level 2: Inductive Refinement

Inductive refinement is motivated by transferring ideas of concept formation to perceptual category learning. By comparing different sketches of objects, which should fall under the same category, the system should be able to construct a description of this category in terms of the relevant visual features. The inductive refinement proposed here combines a generalization of classified sketches as well as a clustering of subsets of the classified objects.

Figure 3 illustrates an example: four sketches of stoves are compared. All of them have a cubic shape and share significant elements of stoves such as hot-

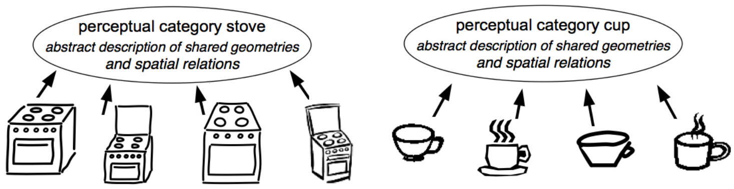

Figure 3. A structural comparison of sketches reveals commonalities that all sketches share.

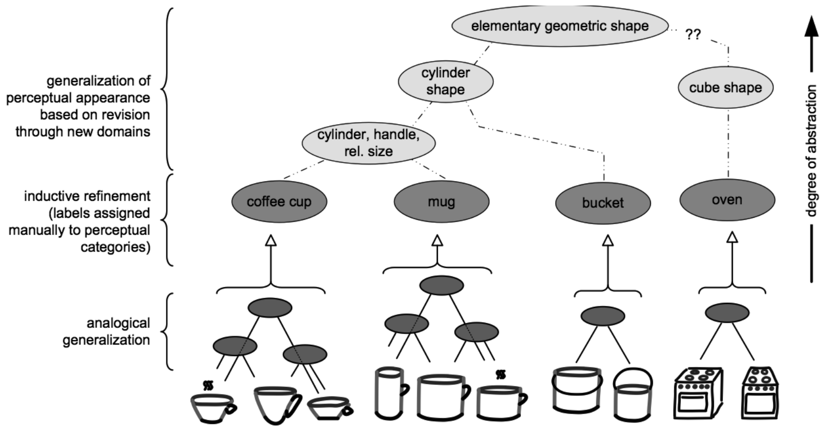

Figure 4. Hierarchical structure of categories learned from sketches.

plates and temperature regulators. Given a pair of sketches, the first level of the proposed system detects the analogous structure and constructs a generalization containing the commonalities as a by-product. This generalization represents the first step towards the perceptual category stove at an abstract level. Iterating this process with additional input stimuli and computing generalizations of already computed generalization candidates will elaborate this category. More generally, provided a set of sketches is given, the exemplified brute force approach would be to compute for each pair of sketches a generalization. These generalizations function as candidates for new perceptual categories, and can be ordered according to their generalization complexity (e.g. substitution lengths in HDTP: The smaller the substitution lengths in the anti-unification process, the more plausible it is to assume that the two input sketches belong to the generated candidate for a perceptual category). The ordered set of candidate generalizations can be used for further structural comparison via anti-unification in order to find commonalities between more than two sketches. Applying clustering techniques may possibly identify optional elements of sketches that appear in many but not all objects (e.g. water vapor over the cups in Figure 3).

3.3. Level 3: Creating a Perceptual Category Hierarchy

Analogies are not only iteratively applied among instances of the same category (drawings of cups), but also between sketch drawings of different categories such as cups, mugs, buckets etc., so that a hierarchy of perceptual categories is attempted to be built (cf. Figure 4). Generated perceptual categories from Level 2 will constitute the leaves of the hierarchy. By analogical comparison of pairs of perceptual categories, generalizations are computed that can represent candidates of new, more abstract perceptual categories. These candidates can be ordered according to the complexity of the underlying analogical mapping and only those candidates constitute new categories that are maximally similar to each other. The generalizations successively reach an abstraction level such that the highest level of generalizations contains elementary geometric shapes.

3.4. The Recognition Task

The recognition task refers to the problem of determining whether a given sketch corresponds to an object from the system's knowledge base. It can also be treated as an analogy problem, in which the source domain consists of the system's knowledge on how to sketch a certain object, and has to be mapped to the unstructured graphical input (target) presented to the system as a flat collection of lines and dots. The structural commonality between the flat representation of the target and the structured representation of the source is initially not obvious. To successfully classify a new stimulus, an analogous structure has to be created for the target stimulus. During the analogy-based mapping process the target must be re-represented such that common structures may become visible.

The hierarchical memory structure built by the system (cf. Figure 4) is used as a starting point for the retrieval. The search algorithm will try to map abstract categories from that hierarchy to the search item, by computing appropriate substitutions to prove that the search item is a suitable instance of that abstract category. Hence, the retrieval is organized as a top-down search: starting from the most abstract category, all sub-categories are analogically mapped to the query sketch. Good matches are those categories where the aligned elements reach maximal coverage of the stored descriptions as well as maximal coverage of the search item. Matching items are all those sketches which are classified below a suitable category in the hierarchy. Suitable categories need to exhibit a sufficiently high coverage of the search item and the category itself. The result of a retrieval process ranks all matching items according to their relevance. We suggest the following criteria to determine the degree of relevance:

1. Depth of the matching database category: The higher a matching category in the hierarchy, the more abstract it is.
2. Coverage of the analogy: We assume that the higher the coverage of the search item, the better is the match.
3. The analogical relation between the search and the database items should be a coherent and connected match. This indicates that not only single elements align, but at least a certain part of the sketch aligns coherently.

In a ranking heuristics that combines the different aspects, the coverage has to be considered with respect to the abstractness of the database category.

4. Related Work

The ideas presented here build on two research fields: spatial analogies and category learning with analogies. Spatial analogies have a rather long history in artificial intelligence, whereas analogy-based learning is far less developed. The first analogy system, ANALOGY [19], was dedicated to solving proportional geometric analogy problems. O'Hara & Indurkhya [20,21] proposed InterAct, an algebraic analogy model for geometric proportional analogies between line drawings. Dastani [22] developed a formal language for this analogy model to describe elements in geometric figures and compute automatically a structural, Gestalt-based representation. Forbus et al. [11] developed a general architecture for sketch understanding, CogSketch, which is domain independent and takes freehand sketches as input [23]. Each freehand sketch drawing consists of several primitive elements called glyphs. CogSketch interprets the primitive elements via their ontological description and via their shapes, and computes spatial relations between primitive elements based on the convex hull of glyphs. Copycat is a non-deterministic analogy model for proportional analogies in the string domain [24]. Tabletop [25] is a computational program based on Copycat that was developed to detect analogous spatial arrangements in a micro-world such as a well-laid table. Like Copycat, Tabletop combines representation-building and correspondence-finding into one integrated process. Davies and colleagues examine visual analogies in architectural design. They showed in experiments [26] that humans use visuospatial representations for the analogical mapping and transfer: participants used visual and spatial knowledge, mostly the topology of objects, to align a given architectural design with an architectural design problem and construct a solution via analogical transfer. Davies et al. developed the analogy model Galatea, an implementation of the constructive adaptive visual analogy theory [27,28], to compute visuospatial analogies.

Analogy-based learning differs from the enormous number of proposed classes of learning methods and methodologies in classical artificial intelligence research, as for example, instance-based learning, exemplar-based learning, case-based learning in the area of lazy learning and version space learning, decision tree learning, inductive learning, neural learning, and probabilistic learning in the area of eager learning. Many of these approaches require a relatively large sample of examples in order to learn reasonable generalizations. Although there may be certain approaches that attempt to incorporate structure of the generalization space into the learning process, in order to facilitate learning from small training data samples – similar to analogical learning – there are significant differences between these approaches and analogy-based learning. Only a rather limited number of positive examples are required for learning due to the conceptually guided way of establishing analogical generalizations, which are the source for new knowledge. An explicit generalization is necessary to capture new categories, re-use learned knowledge, and refine knowledge over learning steps. It is worth pointing out that one can find quite often references to analogical learning [29], but no spelled-out theory of analogical learning has been proposed so far. Inductive Logic Programming (ILP) [30] and Relational Learning [31] could be mentioned as a modern probabilistic version of frameworks where structure plays an important role in the learning approach. But compared to these most prominent approaches, the computation of an analogical relation does not incorporate probabilities, nor does it require that examples are taken from the same domain. However, the computation of an analogical relation is a complex process including aspects like retrieval, transfer, re-representation, refinement etc. Closest in spirit to analogy-making, may be the approach originally proposed by Plotkin [17], who computed least general generalizations for facilitating learning.

5. Summary and Future Work

We have outlined ideas for a system to model sketch learning and recognition. The setup is motivated by psychological findings emphasizing that human recognition capabilities are not only data-driven, but crucially governed by cognitive mechanisms and principles such as analogical reasoning and Gestalt principles. This contrasts with most work in the context of image retrieval, which use low-level features and does not guarantee that the resulting model reflects the human competence in recognition processes, as many of the used features are possibly not accessible by humans. One of the rare exceptions is [32] who propose to view image retrieval as a knowledge representation problem, where structured objects are retrieved such that syntactic and semantic aspects play an important role.

Even though the work presented here is currently purely conceptual, we have explained in detail how the envisaged system can make use of existing technologies, especially from the field of spatial and analogical reasoning. We have argued in favour of a symbolic representation of visual scenes and have proposed to use HDTP as a framework for analogy making. For our system, HDTP has to be extended to make use of spatial reasoners, e.g. from the SparQ toolbox [13], for re-representation during the analogical mapping. A prototype implementation may be applied to a set of test sketches, allowing to compare different heuristics. A primary concern is the development of a suitable language for describing shapes and sketches. Here we can build on a plethora of existing semiformal and formal approaches, like Dastani's languages of perception [22]. Central objectives for such a language are, that it allows for cognitively plausible representation and supports the manipulations required by our system.

References

[1] C. Freksa. Qualitative spatial reasoning. In *Cognitive and Linguistic Aspects of Geographic Space*, pages 361–372. Kluwer Academic Publishers, Dordrecht, 1991.

[2] B. Kuipers. Modeling spatial knowledge. *Cognitive Science: A Multidisciplinary Journal*, 2(2):129–153, 1978.

[3] K. Lynch. *The Image of a City.* MIT Press, Cambridge, MA, 1960.

[4] D.R. Montello. Spatial cognition. In N.J. Smelser and P.B. Baltes, editors, *International Encyclopedia of the Social & Behavioral Sciences*, pages 14771–14775. Pergamon Press, Oxford, 2001.

[5] M. Wertheimer. Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie*, 61(1):161–265, 1912.

[6] M. Wertheimer. *Productive Thinking.* Harper & Row, New York, 1954.

[7] K. Koffka. *Principles of Gestalt Psychology.* Harcourt, New York, 1935.

[8] W. Köhler. *Gestalt Psychology.* Liveright, New York, 1929.

[9] D. Marr. *Vision.* W. H. Freeman and Company, New York, 14th ed. edition, 2000.

[10] I. Biedermann. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2):115–147, 1987.

[11] K.D. Forbus, J. Usher, A. Lovett, K. Lockwood, and J. Wetzel. CogSketch: Open-domain sketch understanding for cognitive science research and for education. In *EUROGRAPHICS Workshop on Sketch-Based Interfaces and Modeling*, 2008.

[12] B. Indurkhya. Modes of analogy. In K.P. Jantke, editor, *Analogical and Inductive Inference*, number 397 in LNAI, pages 217–230, 1989.

[13] D. Wolter. SparQ - a spatial reasoning toolbox. In *AAAI Spring Symposium on Benchmarking of Qualitative Spatial and Temporal Reasoning Systems*, 2009.

[14] M. Westphal, S. Wöfl, and Z. Gantner. Gqr: A fast solver for binary qualitative constraint networks. In *AAAI Spring Symposium on Benchmarking of Qualitative Spatial and Temporal Reasoning Systems*, 2009.

[15] H. Gust, U. Krumnack, K.-U. Kühnberger, and A. Schwering. Integrating analogical and inductive learning at different levels of generalization. In *Workshop on Learning from Non- Vectorial Data (LNVD2007) at the 30th Annual German Conference on Artificial Intelligence (KI07)*, pages 46–57, 2007.

[16] A. Schwering, U. Krumnack, K.-U. Kühnberger, and H. Gust. Syntactic principles of Heuristic-Driven Theory Projection. *Special Issue on Analogies - Integrating Cognitive Abilities. In: J. of CSR*, 10(3):251–269, 2009.

[17] G. D. Plotkin. A note on inductive generalization. *Machine Intelligence*, 5:153–163, 1969.

[18] U. Krumnack, H. Gust, K.-U. Kühnberger, and A. Schwering. Re-representation in a logic-based model for analogy making. In Wayne R. Wobcke and Mengjie Zhang, editors, *AI 2008: Advances in AI, 21st Australasian AI Conference (AI-08)*, volume 5360 of *LNAI*, pages 42–48. Springer, 2008.

[19] T. G. Evans. A heuristic program to solve geometric-analogy problems. In *Proc. of the April 21-23, 1964, spring joint computer conference*, pages 327–338, 1964.

[20] S. O'Hara. A model of the redescription process in the context of geometric proportional analogy problems. In *Proc. of the International Workshop on Analogy and Inductive Inference (AII'92)*, pages 268–293. Springer, 1992.

[21] S. O'Hara and B. Indurkhya. Incorporating (re)-interpretation in case-based reasoning. In *Topics in Case-Based Reasoning*, volume 837 of *LNCS*, pages 246–260. Springer, Berlin, Heidelberg, 1994.

[22] M.M. Dastani. *Languages of Perception*. PhD thesis, University of Amsterdam, 1998.

[23] K.D. Forbus, K. Lockwood, M. Klenk, E. Tomai, and J. Usher. Open-domain sketch understanding: The nuSketch approach. In *AAAI Fall Symposium on Making Pen-based Interaction Intelligent and Natural*, Washington, DC, 2004.

[24] D. R. Hofstadter and M. Mitchell. The copycat project: A model of mental fluidity and analogy-making. In K. Holyoak and J. Barnden, editors, *Advances in Connectionist and Neural Computation Theory Vol 2: Analogical Connections*, pages 31–112. Ablex Publishing Corporation, Norwood NJ, 1994.

[25] R. M. French. The computational modeling of analogy-making. *Trends in Cognitive Science*, 6(5):200–205, 2002.

[26] C. Davies, A.K. Goel, and N.J. Nersessian. A computational model of visual analogies in design. *CSR*, 10(3):204–215, 2009. Special Issue on Analogies – Integrating Cognitive Abilities.

[27] C. Davies. *Constructive adaptive visual analogy*. PhD thesis, College of Computing,, Georgia Institute of Technology: Georgia, 2004.

[28] J. Davies and A.K. Goel. Representation issues in visual analogy. In *Proc. of the 25th Annual Conference of the CogSci Society*, 2003.

[29] Dedre Gentner. The mechanisms of analogical learning. In S. Vosniadou and A. Ortony, editors, *Similarity and analogical reasoning*, pages 197–241. Cambridge University Press, Cambridge, 1989.

[30] S.H. Muggleton. Inductive logic programming. *New Generation Computing*, 8(4):295–318, 1991.

[31] L. de Raedt. *Logical and Relational Learning*. Cognitive Technologies. Springer, Berlin, Heidelberg, 2008.

[32] Eugenio Di Sciascio, F. M. Donini, and M. Mongiello. Spatial layout representation for query-by-sketch content-based image retrieval. *Pattern Recognition Letters*, 23:1599–1612, 2002.