08091 Abstracts Collection
Logic and Probability for Scene Interpretation
— Dagstuhl Seminar —

Bernd Neumann[1], Anthony G. Cohn[2], David C. Hogg[3] and Ralf Möller[4]

[1] University of Hamburg, DE
neumann@informatik.uni-hamburg.de
[2] Leeds University, GB
a.g.cohn@leeds.ac.uk
[3] Leeds University, GB
d.c.hogg@leeds.ac.uk
[4] Hamburg University of Technology, DE
r.f.moeller@tu-harburg.de

Abstract. From 25.2.2008 to Friday 29.2.2008, the Dagstuhl Seminar 08091 "Logic and Probability for Scene Interpretation" was held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper.

Keywords. Logic, probabilities, scene interpretation

08091 Executive Summary – Logic and Probability for Scene Interpretation

From 25.2.2008 to Friday 29.2.2008, the Dagstuhl Seminar 08091 "Logic and Probability for Scene Interpretation" was held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. Links to extended abstracts or full papers are provided, if available.

The program consisted of 21 talks and discussions, attended by 40 participants from 12 countries. In accordance with the interdisciplinary nature of the workshop topic, the attendants represented distinct streams in Computer Vision, AI and Cognitive Science, in particular High-level Computer Vision, Logical Models in AI, Probabilistic Models in AI, Robotics, Multimedia Content Representation, and Cognitive Models.

The main goal of the workshop "to advance the use of logic-based knowledge representation and reasoning for scene interpretation (static and dynamic

scenes), and explore possible ways for reconciling logics with probabilistic models" has been reached in several respects.

1. Several contributions showed the power but also the complexity of logical models for scene interpretation:
   - Brandon Bennett, Leeds University Enhancing Tracking by Enforcing Spatio-Termporal Consistency Constraints
   - Francois Bremond, INRIA, Sophia Antipolis Temporal scenarios for automatic video interpretation
   - Britta Hummel, Karlsruhe University Description Logic for Intersection Understanding
   - Ralf Möller, Technical University Hamburg-Harburg A Logical Model for Multimedia Interpretation
   - Fiora Pirri, University of Roma "La Sapienza" Lifting Models for Scene Interpretation
   - Matthias Schlemmer, Technical University Vienna Abstraction, ontology and task-guidance for visual perception in robots
2. Advanced combinations of probabilistic and logical models for scene interpretation were presented by
   - Bastian Leibe, ETH Zürich Mobile Scene Understanding Integrating Recognition, Reconstruction, and Tracking
   - Dima Damen, Leeds University Constrained Scene Interpretation - Sequences of Uncertain Events
   - Otthein Herzog, Bremen University Qualitative Abstraction and Inherent Uncertainty in Scene Recognition
   - Pascal Hitzler, Karlsruhe University Approximate Reasoning with OWL Ontologies
   - Manfred Jaeger, Aalborg UniversityCombining probabilistic graphical models and logic: the Markov Logic and Relational Bayesian Network approaches
   - Hans-Hellmut Nagel, Karlsruhe University Toward Algorithmic Generation of Temporal-logic Representations for Driver Behavior from Legal Texts
   - Bernd Neumann, Hamburg University Probabilistic Inferences in Compositional Hierarchies
   - Maria Petrou, Imperial College, London Tower of Knowledge
3. The need for learning proved to be an important motivation for employing probabilistic models. Several contributions addressed learning aspects of scene interpretation:
   - Meg Aycinena, MIT, Cambridge Learning Grammatical Models for Object Recognition
   - Luc De Raedt, Freiburg University Statistical Relational Learning - a Logical Introduction
   - Paulo Santos, Centro Universitario di FEI, Sao Paulo Assimilating knowledge from neuroimages in schizophrenia diagnostics

   - Sven Wachsmuth and Agnes Swadzba, Bielefeld University Probabilistic modeling of spatial and temporal aspects of scenes for human-robot interaction

4. Cognitive aspects turned out to be less widely discussed. Many thanks go to Aaron Sloman who emphasized these apects in discussions and in his talk:
   - Aaron Sloman, The University of Birmingham What are we trying to do, and how do logic and probability fit into the bigger picture?The participants were also reminded of the power of pattern-based approaches by the talks:
   - Vaclav Hlavac, Czech Technical University, Prague Image and Structure
   - Diedrich Wolter, Bremen University Qualitative Arrangement Information for Matching

In the final discussion, there was overwhelming agreement regarding the usefulness of the exchange of ideas and results in this workshop in general, and also regarding several insights in particular:

- A necessary third topic besides logics and probabilities in scene interpretation is learning. Scene interpretation appears to be an excellent topic for the learning community.
- There is urgent need for benchmarks, data sets and videos to be shared within the interdisciplinary scene interpretation community.

However, unanswered questions and unsolved problems by far outnumbered definite answers to the questions with which the workshop has started. Here are some of these open questions:Are there a priori roles for logics and probabilities in scene interpretation?What is the proper semantics for peobabilities in scene inetrpretation?What is the impact of high-level and common sense knowledge in scene interpretation?What are the consequences of logics and probabilities for a scene interpretation system architecture?Thanks to the professional and friendly people of the Dagstuhl organisation! Many of the participants hope to convene again in Dagstuhl for discussions of a similar topic.

Learning Grammatical Models for Object Recognition

Meg Aycinena (MIT - Cambridge, US)

Many object recognition systems are limited by their inability to share common parts or structure among related object classes. This capability is desirable because it allows information about parts and relationships in one object class to be generalized to other classes for which it is relevant. This ability has the potential to allow effective parameter learning from fewer examples and better generalization of the learned models to unseen instances, and it enables more efficient recognition. With this goal in mind, we have designed a representation and recognition framework that captures structural variability and shared part

structure within and among object classes. The framework uses probabilistic geometric grammars (PGGs) to represent object classes recursively in terms of their parts, thereby exploiting the hierarchical and substitutive structure inherent to many types of objects. To incorporate geometric and appearance information, we extend traditional probabilistic context-free grammars to represent distributions over the relative geometric characteristics of object parts as well as the appearance of primitive parts. We describe an efficient dynamic programming algorithm for object categorization and localization in images given a PGG model. We also develop an EM algorithm to estimate the parameters of a grammar structure from training data, and a search-based structure learning approach that finds a compact grammar to explain the image data while sharing substructure among classes. Finally, we describe a set of experiments that demonstrate empirically that the system provides a performance benefit.

Keywords: Object recognition, grammars, structure learning

Joint work of: Aycinena, Meg; Kaelbling, Leslie Pack; Lozano-Perez, Tomas

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1611

Combining Logic and Probability

Brandon Bennett (University of Leeds, GB)

The paper gives a high-level overview of some ways in which logical representations and reasoning can be used in computer vision applications, such as tracking and scene interpretation. The combination of logical and statistical approaches is also considered.

Keywords: Vision, Tracking, Logic, Probability, Spatio-Temporal Continuity

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1612

Temporal scenarios for automatic video interpretation

Francois Bremond (INRIA - Sophia Antipolis, FR)

I have presented our work on Temporal Scenario Representation and Recognition for Automatic Video Interpretation.

Our goal is to design an algorithm recognizing in real-time temporal scenarios predefined by experts and taking as input mobile objects tracked by a vision module and a priori knowledge of the observed environment. We have proposed an approach addressing two issues: scenario representation and scenario recognition.

Concerning the issue of temporal scenario representation: we have proposed a video event ontology (in collaboration with an ARDA workshop series on video

events) that can facilitate the representation of temporal scenarios. This ontology is composed of concepts to describe physical objects (e.g. mobile-objects, contextual-objects), video events (e.g. primitive-state, composite-event) and relations between concepts (e.g. temporal-relations, spatial-relations). Then, based on this ontology, we have proposed a description language helping experts of different domains to describe easily their scenarios of interest. The ontology and the language are/have been used by experts of four European/French projects for video surveillance (CASSIOPEE for bank agency surveillance, AVITRACK for apron monitoring, SAMSIT for inside train surveillance and ADVISOR for metro station surveillance).

Concerning the recognition issue, a challenge is to reduce the complexity of the temporal scenario recognition algorithm. The recognition can be viewed as a Temporal Constraint Satisfaction Problem (TCSP). So, the algorithm to solve this problem belongs generally to the NP-complete class. Thus, the other tasks of our work concern the reduction of the processing time for the recognition algorithm. To solve this problem, we have proposed a temporal scenario recognition method that is able to recognize predefined scenarios in real-time (in video cadence). We have extend this method to handle (1) audio-video data and (2) scenario uncertainty.

To validate the recognition algorithm, we have cooperated with the experts of four projects for video surveillance/monitoring to realize three types of tests with: recorded videos, live videos and simulated data. The obtained results answer requirements. The recognition algorithm together with the vision routines can recognize correctly in real-time (10 frames/second) scenario occurrences in a longtime interval (a week). Moreover, the recognition algorithm can cope with complex videos containing 240 persons and with complex scenarios defined with 10 physical objects and 10 components.

Keywords: Video interpretation, temporal scenarios

Joint work of: Bremond, Francois; Thinh, VuVan; Monique, Thonnat

Probabilistic Logic Learning: An Introduction

Luc De Raedt (Katholieke Universiteit Leuven, BE)

Statistical Relational Learning is a new subfield of artificial intelligence lying at the intersection of machine learning, reasoning about undertainty and relational and logical representations. It aims at developing models that can elegantly deal with objects as well as the relationships that hold amongst them. In this talk, I shall first motivate this research stream using a number of applications and then analyse its state-of-the-art taking a logical perspective.

More specifically, I shall explore the relationships between the traditional probabilistic models, which work essentially with a propositional representation, and their upgrades within statistical relational learning and introduce these using a number of examples. The examples used will include the upgrading of

Bayesian or Markov networks (towards PRMs, BLPs or MLNs) and those of probabilistic Context Free Grammars and HMMs (towards SLPs, PRISM, ICL and LOHMMs) as well as our recent work on ProbLog and link mining in large biological networks.

Keywords: Probability, logic, learning, inductive logic programming, statistical relational learning

Qualitative Abstraction and Inherent Uncertainty in Scene Recognition

Otthein Herzog (Universität Bremen, DE)

The interpretation of scenes, e.g., in videos, is demanding at all levels. At the image processing level it is necessary to apply an "intelligent" segmentation and to determine the objects of interest. For the higher symbolic levels it is a challenging task to perform the transition between quantitative and qualitative data and to determine the relations between objects. Here we assume that the position of objects ("agents") in images and videos will already be determined as a minimal requirement for the further analysis.

The interpretation of complex and dynamic scenes with embedded intentional agents is one of the most challenging tasks in current AI and imposes highly heterogeneous requirements. A key problem is the efficient and robust representation of uncertainty. We propose that uncertainty should be distinguished with respect to two different epistemological sources: (1) noisy sensor information and (2) ignorance. In this presentation we propose possible solutions to this class of problems.

The use and evaluation of sensory information in the field of robotics shows impressive results especially in the fields of localization (e.g. MCL) and map building (e.g. SLAM) but also imposes serious problems on the successive higher levels of processing due to the probabilistic nature. In this presentation we propose that the use of (a) qualitative abstraction (classic approach) from quantitative to (at least partial) qualitative representations and (b) coherence-based perception validation based on Dempster-Shafer (DST) can help to reduce the problem significantly.

The second important probability problem class that will be addressed is ignorance. In our presentation we will focus on reducing missing information by inference. We contrast/compare our experiences in an important field of scene interpretation namely plan and intention recognition. The first approach is based on a logical abductive approach and the second approach in contrast uses a probabilistic approach (Relational Hidden Markov Model (RHMM)).

Keywords: Scene interpretation, intentional agents, uncertainty, qualitative abstraction, coherence-based perception, abduction, RHMM

Joint work of: Elfers, Carsten; Herzog, Otthein; Miene, Andrea; Wagner, Thomas

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1614

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2008/1614

Approximate OWL Instance Retrieval with SCREECH

Pascal Hitzler (Universität Karlsruhe (TH), DE)

With the increasing interest in expressive ontologies for the Semantic Web, it is critical to develop scalable and efficient ontology reasoning techniques that can properly cope with very high data volumes. For certain application domains, approximate reasoning solutions, which trade soundness or completeness for increased reasoning speed, will help to deal with the high computational complexities which state of the art ontology reasoning tools have to face. In this paper, we present a comprehensive overview of the SCREECH approach to approximate instance retrieval with OWL ontologies, which is based on the KAON2 algorithms, facilitating a compilation of OWL DL TBoxes into Datalog, which is tractable in terms of data complexity. We present three different instantiations of the Screech approach, and report on experiments which show that the gain in efficiency outweighs the number of introduced mistakes in the reasoning process.

Keywords: Description logics, automated reasoning, approximate reasoning, Horn logic

Joint work of: Hitzler, Pascal; Krötzsch, Markus; Rudolph, Sebastian; Tserendorj, Tuvshintur

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1615

Image and Structure

Vaclav Hlavac (Czech Technical University, CZ)

Statistical pattern recognition methods have had difficulties to deal with images for several decades because of the structure induced by the neighborhood relation among pixels and spatial arrangements among image entities on the larger scale level. Many researchers believe that attempts to apply structural pattern recognition methods in the 1960s and 1970s and logical reasoning pursuit in artificial inteligence led to the dead end.

I like to advocate that the structural pattern recognition can be embedded into the statistical pattern recognition framework. This step has a potential to bring robustness to the structural approach as well as the ability to cope with noise. The talk will first explain the theoretical anchoring of the approach and second it will give a few examples from our recent research will, e.g., (a) optimizations on Markovian random fields applied to non-rigid matching in images or segmentation; (b) the structural construction applied to on-line grammar-based recognition of mathematical formulae.

Keywords: Structural pattern recognition, optimization on Markov Random Fields

Joint work of: Hlavac, Vaclav; Shekhovtsov Alexander; Werner, Tomas

Scene Understanding of Urban Road Intersections with Description Logic

Britta Hummel (Universität Karlsruhe, DE)

Road recognition from video sequences has been solved robustly only for small, often simplified subsets of possible road configurations. A massive augmentation of the amount of prior knowledge may pave the way towards a generation of estimators of more general applicability. This contribution introduces Description Logic extended by rules as a promising knowledge representation formalism for road and intersection understanding.

We have set up a Description Logic knowledge base for arbitrary road and intersection geometries and configurations. Logically stated geometric constraints and road building regulations constrain the hypothesis space. Sensor data from an in-vehicle vision sensor and from a digital map provide evidence for a particular intersection. Partial observability and different abstraction layers of the input data are naturally handled by the representation formalism.

Deductive inference services - namely satisfiability, classification, entailment, and consistency - are then used to narrow down the intersection hypothesis space based on the evidence and the background knowledge, and to retrieve intersection information relevant to a user, i.e. a human or a driver assistance system. We conclude with an outlook towards non-deductive reasoning, namely model construction under the answer set semantics.

Keywords: Autonomous Driving, Road Recognition, Knowledge Representation, Description Logic, Nonmonotonic Reasoning

Joint work of: Hummel, Britta; Thiemann, Werner; Lulcheva, Irina

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1616

See also: B. Hummel, W. Thiemann, I. Lulcheva. Description Logic for Vision-Based Intersection Understanding. In Proc. Cognitive Systems with Interactive Sensors (COGIS), Stanford University, CA, 2007

Combining probabilistic graphical models and logic

Manfred Jaeger (Aalborg University, DK)

I will review and compare two approaches that combine elements of predicate logic with probabilistic graphical models: Markov Logic Networks, and Relational Bayesian Networks.

While both approaches are based on similar motivations, they follow rather different representation strategies, being based on undirected and directed graphical models, respectively.

I will propose a general semantic framework for the analysis of these (and other, related) models, and show how Markov Logic Networks can be encoded by Relational Bayesian Networks.

Robust Multi-Person Tracking from Moving Platforms

Bastian Leibe (ETH Zentrum - Zürich, CH)

In this paper, we address the problem of multi-person tracking in busy pedestrian zones, using a stereo rig mounted on a mobile platform. The complexity of the problem calls for an integrated solution, which extracts as much visual information as possible and combines it through cognitive feedback. We propose such an approach, which jointly estimates camera position, stereo depth, object detection, and tracking. We model the interplay between these components using a graphical model. Since the model has to incorporate object-object interactions, and temporal links to past frames, direct inference is intractable. We therefore propose a two-stage procedure: for each frame we first solve a simplified version of the model (disregarding interactions and temporal continuity) to estimate the scene geometry and an overcomplete set of object detections. Conditioned on these results, we then address object interactions, tracking, and prediction in a second step. The approach is experimentally evaluated on several long and difficult video sequences from busy inner-city locations. Our results show that the proposed integration makes it possible to deliver stable tracking performance in scenes of realistic complexity.

Keywords: Pedestrian detection, tracking, Mobile vision

Joint work of: Ess, Andreas; Schindler, Konrad; Leibe, Bastian; van Gool, Luc

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1617

A Logical Model for Multimedia Interpretation

Ralf Möller (TU Hamburg-Harburg, DE)

In order to organize the retrieval of multimedia documents, in many applications, keywords are used but are usually found to be less than optimal for a number of reasons. Furthermore, modalities such as images or audio and video are known to be hard to index using string-based access techniques. Manual annotation with keywords and labels is the state-of-the-art in publishing companies. Although one might think about keywords as being concept names arranged a hierarchy (e.g., via ontologies), and thus, query answering might exploit abstraction and

synonyms, the subjectiveness of natural language annotation is one the main reasons for problems in information retrieval processes. In addition to concept labels, It might also be important to refer to specific parts and their relations to best support dedicated information retrieval requirements. Acquiring such deep-level descriptions of multimedia documents in a manual way is often too expensive, and even a coarse description that is automatically generated will be very helpful for many information retrieval tasks. For instance, considering images in an athletics domain, we can see that from detectable objects such as a crossbar, a person and a pole (together with respective spatial relations), a more high-level event such as pole vault might be postulated (and used for supporting retrieval processes later on).

In order to support the automatic derivation of metadata as descritions for multimedia documents we developed an abduction-based approach which builds upon ontologies (description logics). Given basic objects and their relations detectable via (low-level) analysis processes, the abduction approach computes explanations in terms of instance descriptions (Aboxes) such that detected relations (or assertions) for detected objects are entailed by the explanations and the background ontology (Tbox). The explanations that are consistent for all analysis results comprise a so-called media interpretation. Logic is used to define the space of possible interpretations. In particular, rules and Tbox axioms are used to define the space of abducibles. Obviously, since there might multiple explanations possible w.r.t. larger ontologies, there might be multiple interpretations for certain parts of a media document.

Although even for a crisp logic we were able to define a preference measure on possible interpretations, in quite a number of cases even after fusing the interpretation results w.r.t. several modalities (e.g., image and text on a web page) too many alternative interpretations remain. We therefore investigated probabilistic approaches based on the view that high-level objects are seen as aggreates. The preference measure uses probabilities on instance descriptions (Aboxes) as well as probabilistic ontological knowledge (Tboxes) to rank alternative media interpretation Aboxes.

In the workshop presentation we will analyse the requirements for a combination of logic and probabilities for deriving coarse high-level descriptions of multimedia documents to be used in content-based information retrieval.

Keywords: Media interpretation, abduction, probabilistic ranking

Joint work of: Möller, Ralf; Kaya, Atila; Näth, Tobias; Espinosa Peraldi, Sofia

See also: S. Espinosa Peraldi, A. Kaya, S. Melzer, R. Möller, and M. Wessel. Towards a Media Interpretation Framework for the Semantic Web. The 2007 IEEE/WIC/ACM International Conference on Web Intelligence (WI'07), 2007.

Toward Algorithmic Generation of Temporal-logic Representations for Driver Behavior from Legal Texts

Hans-Hellmut Nagel (Universität Karlsruhe, DE)

A comprehensive Driver Assistance System (DAS) is expected to advise drivers in conformance with traffic laws and regulations.

Such an expectation implies that a DAS is capable to take the semantics of legal texts into account when formulating its advice to drivers.

The semantic content of traffic laws and regulations thus must be accessible to DAS inference processes activated during (i) the evaluation of sensory - in particular video - signals, (ii) the extraction of conclusions regarding the optimal (re-)action to the currently sensed traffic situation, and during (iii) the ensueing Natural Language Advice Generation. Obviously, a system with such capabilities can be used, too, in order to report temporal developments in traffic scenes recorded by one or more video camera(s) in the form of natural language texts. Moreover, reliable extraction and assessment of traffic situations based on video recordings facilitates the introduction of video cameras into the feedback loop of DAS maneuvers whose execution has been authorized by the driver. Traffic laws and regulations are already expressed in a highly refined semi-formal language ('legalese'), due to about a century of efforts by lawyers and judges. It thus appears sensible to treat the legal texts as a formal specification of law-conforming driver behavior. Transforming this body of texts into the algorithms of a DAS is a time-consuming and potentially error-prone process. One possibility could be to create a logic process which verifies the conformity of DAS-(re-)actions with the legal texts. Alternatively, one may develop a direct algorithmic transformation of relevant legal texts into a temporal-logic representation suitable for an appropriate inference machine. It then is the task of such an inference machine to instantiate generic rules based on the evaluation of current sensory signals and to convert the instantiated rules into law-conforming recommendations to the driver. Both approaches necessitate a careful formalisation of common sense knowledge regarding road traffic. Given the complexity and challenge implied by the latter alternative, investigations into this direction appear recommendable. An explorative investigation along the latter path has been started based on temporal logic for behavior representation and a related inference engine, in combination with video evaluation processes and a Natural Language Text Generation component.

Experiences with and (currently perceived) bottlenecks of this approach are discussed.

Keywords: Behavior Representation, Fuzzy Metric-Temporal Logic, Text Understanding, Text-to-Logic, Logic-to-Actions

Bayesian Compositional Hierarchies - A Probabilistic Structure for Scene Interpretation

Bernd Neumann (Universität Hamburg, DE)

In high-level vision, it is often useful to organize conceptual models in compositional hierarchies. For example, models of building facades (which are used here as examples) can be described in terms of constituent parts such as balconies or window arrays which in turn may be further decomposed. While compositional hierarchies are widely used in scene interpretation, it is not clear how to model and exploit probabilistic dependencies which may exist within and between aggregates. In this contribution I present Bayesian Aggregate Hierarchies as a means to capture probabilistic dependencies in a compositional hierarchy. The formalism integrates well with object-centered representations and extends Bayesian Networks by allowing arbitrary probabilistic dependencies within aggregates. To obtain efficient inference procedures, the aggregate structure must possess abstraction properties which ensure that internal aggregate properties are only affected in accordance with the hierarchical structure. Using examples from the building domain, it is shown that probabilistic aggregate information can thus be integrated into a logic-based scene interpretation system and provide a preference measure for interpretation steps.

Keywords: Scene interpretation, compositional hierarchy, probabilistic inference

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1605

The Tower of Knowledge: a novel architecture for organising knowledge combining logic and probability

Maria Petrou (Imperial College London, GB)

It is argued that the ability to generalise is the most important characteristic of learning and that generalisation may be achieved only if pattern recognition systems learn the rules of meta-knowledge rather than the labels of objects. A structure, called "tower of knowledge", according to which knowledge may be organised, is proposed. A scheme of interpreting scenes using the tower of knowledge and aspects of utility theory is also proposed. Finally, it is argued that globally consistent solutions of labellings are neither possible, nor desirable for an artificial cognitive system.

Keywords: Learning by example, learning rules

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1606

See also: Progress in Pattern Recognition, Image Analysis and Applications. The 12th Iberoamerican Congress on Pattern Recognition, CIARP 2007, Vina del Mar-Valparaiso, November, L Rueda, D Mery and J Kittler (eds), LNCS 4756, Springer, pp 1-12.

lifting models for scene interpretation

Fiora Pirri (University of Rome "La Sapienza", IT)

Relevant neurological and psychological studies in early perception and results obtained in the computational theory of perception such as in cognitive vision, robotic manipulation, artificial sniffing, mapping and localization, are mostly based on mathematical and statistical models. These models are used to both explain and simulate the behavior of receptive fields and the complex structure of their connections. On the other hand, relevant results on commonsense reasoning and reasoning about actions and the interaction with the environment have provided a rich literature on logical models structuring knowledge concerning spatial reasoning and inference in action domains .

The lack of a pathway between the two classes of models has to be overcome, despite no hint is still given in neurological studies (e.g. the NCC account of consciousness). Even if it is yet not clear how to lift statistically based structures of early perception to logically based models of spatial reasoning the idea of structure transformation, could be more feasible than the idea of just making more and more expressive logical languages. In this presentation I would like to analyse a case of structure transformation from a statistical model to a logical one

Assimilating knowledge from neuroimages in schizophrenia diagnostics

Paulo E. Santos (Centro Universitario da FEI - Sao Paolo, BR)

The aim of this article is to propose an integrated framework for classifying and describing patterns of disorders from medical images using a combination of image registration, linear discriminant analysis and region-based ontologies. In a first stage of this endeavour we are going to study and evaluate multivariate statistical methodologies to identify the most discriminating hyperplane separating two populations contained in the input data. This step has, as its major goal, the analysis of all the data simultaneously rather than feature by feature. The second stage of this work includes the development of an ontology whose aim is the assimilation and exploration of the knowledge contained in the results of the previous statistical methods. Automated knowledge discovery from images is the key motivation for the methods to be investigated in this research. We argue that such investigation provides a suitable framework for characterising the high complexity of MR images in schizophrenia.

Keywords: Statistical classification, spatial ontologies

Joint work of: Santos, Paulo; Thomaz, Carlos; Celiberto, Luiz; Duran, Fabio; Gattaz, Wagner; Busatto, Geraldo

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1607

Abstraction, ontology and task-guidance for visual perception in robots

Matthias Schlemmer (TU Wien, AT)

For solving recognition tasks in order to navigate in unknown environments and to manipulate objects, humans seem to use at least the following crucial capabilities: abstraction (for storing higher-level concepts of things), common sense knowledge and prediction. Whereas the first and second provide the basis for situated recognition, the second and third serve for pruning the search space as it helps anticipating what (in an abstract sense) they will see next and where. The main goal of our current research is, how we could use such a kind of "common sense world knowledge" for guiding visual perception and understanding scenes. Therefore, we are combining an owl-ontology with the output of vision tools. The additional use of abstraction techniques tries to establish the possibility of detecting higher level concepts, such as arches composed of a variable number of parts. The goal is to finally find concepts such as doors and tables in arbitrary scenes in order to arrive at a generic recognition tool for home robots. The ontology should additionally provide task-specific information about the things to detect.

Keywords: Abstraction, ontology, task, vision

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1608

WHAT ARE WE TRYING TO DO, AND HOW DO LOGIC AND PROBABILITY FIT INTO THE BIGGER PICTURE? Understanding the functions of animal vision

Aaron Sloman (University of Birmingham, GB)

As I said when I received the original invitation I don't have expertise regarding *probabilistic* approaches. It seems to me that insofar as manipulation of probabilities has a role in connection with uncertainty due to noise, poor resolution, occlusion, aperture problems, etc. we have no hope of producing good mechanisms unless we have very clear and effective ideas about what needs to be represented when there is NO uncertainty and how that information can be represented, transformed, and used.

Putting in probabilistic mechanisms too soon is like building a repair kit for an engine before you have designed the engine.

As far as the use of logic is concerned, I think that is merely one kind of representation, which is very useful because of its generality, but for many problems involving spatial structures, processes and causal interactions it can be more useful to use spatial (geometric and topological) representations, though not necessarily isomorphic with what they represent – as pointed out in my IJCAI 1971 discussion of the importance of both Fregean and analogical representations, now online here: http://www.cs.bham.ac.uk/research/projects/cogaff/04.html#200407

However it has proved very difficult to design computer based virtual machines with the required properties. Perhaps that is because we are still not clear enough about the requirements. My work is mostly about requirements, but I have some sketchy design ideas.

Keywords: Analogical & Fregean representation animal vision, causation, geometry, processes, proto-affordances, representation, structures, topology

Architectural and Representational Requirements for Seeing Processes, Proto-affordances and Affordances

Aaron Sloman (University of Birmingham, GB)

This paper, combining the standpoints of philosophy and Artificial Intelligence with theoretical psychology, summarises several decades of investigation by the author of the variety of functions of vision in humans and other animals, pointing out that biological evolution has solved many more problems than are normally noticed. For example, the biological functions of human and animal vision are closely related to the ability of humans to do mathematics, including discovering and proving theorems in geometry, topology and arithmetic. Many of the phenomena discovered by psychologists and neuroscientists require sophisticated controlled laboratory settings and specialised measuring equipment, whereas the functions of vision reported here mostly require only careful attention to a wide range of everyday competences that easily go unnoticed. Currently available computer models and neural theories are very far from explaining those functions, so progress in explaining how vision works is more in need of new proposals for explanatory mechanisms than new laboratory data. Systematically formulating the requirements for such mechanisms is not easy. If we start by analysing familiar competences, that can suggest new experiments to clarify precise forms of these competences, how they develop within individuals, which other species have them, and how performance varies according to conditions. This will help to constrain requirements for models purporting to explain how the competences work. For example, Gibson's theory of affordances needs a number of extensions, including allowing affordances to be composed in several ways from lower level proto-affordances. The paper ends with speculations regarding the need for new kinds of infor-mation-processing machinery to account for the phenomena.

Keywords: Vision, affordances, architectures, development, design space

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1656

EXPLOITING SPATIAL CONTEXT IN IMAGE INTERPRETATION USING FUZZY CONSTRAINT REASONING

Steffen Staab (Universität Koblenz-Landau, DE)

We present an approach for integrating explicit knowledge about the spatial context of objects into image region labelling. Our approach is based on spatial prototypes that represent the typical arrangement of objects in images. We use Fuzzy Constraint Satisfaction Problems as the underlying formal model for producing a labelling that is consistent with the spatial constraints of prototypes.

Keywords: Semantic Web, Fuzzy Logics, Segmentation, Annotation

Joint work of: Saathoff, Carsten; Staab, Steffen

Probabilistic Scene Modeling for Situated Computer Vision

Agnes Swadzba, Sven Wachsmuth (Universität Bielefeld, DE)

Verbal statements and vision are a rich source of information in a human-machine interaction scenario. For this reason Situated Computer Vision aims to include knowledge about the communicative situation in which it takes place. This paper presents three approaches how to achieve scene models of such scenarios combining different modalities.

Seeing (planar) scenes as configurations of parts leads to a probabilistic modeling with BayesŠ nets relating spoken utterances with results of an object recognition step. In the second approach parallel datasets form the basis for analyzing the statistical dependencies between them through learning a statistical translation model which maps between these datasets (here: words in a text and boundary fragments extracted in 2D images). The third approach deals with complex indoor scenes from which 3D data is acquired. Planar structures in the 3D points and statistics extracted on these planar patches describe the coarse spatial layouts of different indoor room types in such a way that a holistic classification scheme can be provided.

Keywords: Scene Modeling, Human Robot Interaction

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1609

Qualitative Arrangement Information for Matching

Diedrich Wolter (Universität Bremen, DE)

In the context of a generalized robot localization task we investigate the utility of qualitative arrangement information in recognition tasks. Qualitative information allows us to make certain knowledge explicit, separating it from uncertain information that we are facing in recognition tasks. This can give rise to efficient matching algorithms for recognition tasks. Particularly qualitative ordering information is very helpful: it can adequately capture certain spatial knowledge and leads to efficient polynomial-time matching algorithms.

Keywords: Matching, qualitative spatial reasoning

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1610