Detection and Identification of Rare Audiovisual Cues

Insperata accident magis saepe quam quae speres.
(Things you do not expect happen more often than
things you do expect) Plautus (ca 200(B.C.)

Project no: 027787

DIRAC

Detection and Identification of Rare Audio-visual Cues

Integrated Project
IST - Priority 2
Project no: 027787

DELIVERALBE No: D7.1
First DIRAC Summer School on Cognitive Engineering
[Updated Version]

Date of deliverable: 30.06.2007
Actual submission date: 02.10.2007

Start date of project: 01.01.2006 Duration: 60 months

Organization name of lead contractor for this deliverable:
Katholieke Universiteit Leuven

Revision [1]

Project co-funded by the European Commission within the Sixth Framework Program (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other program participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Insperata accident magis saepe quam quae speres.
(Things you do not expect happen more often than
things you do expect)   Plautus (ca 200(B.C.)

D7.1 FIRST DIRAC SUMMER SCHOOL ON COGNITIVE ENGINEERING [UPDATED VERSION]

Katholieke Universiteit Leuven (KUL)

Abstract:

This deliverable presents the first DIRAC Summer Workshop on Cognitive Engineering. It provides a summary of the organizational efforts as well as a detailed overview of the final program.

Table of Content

1. Introduction

One of the main goals of DIRAC is to provide effective preparation of DIRAC researchers, but also other European researchers, to do research in cognitive engineering. Therefore, a series of summer schools will be established within the DIRAC framework to expose its researchers to the multi-disciplinary nature of the field of cognitive systems.

The first event in this series is a Summer Workshop on Multi-Sensory Modalities in Cognitive Science which was held at the Studienzentrum Gerzensee in Switzerland from 25 till 29 August 2007. It covers a broad range of topics including neural mechanisms of recognition and categorization, visual object recognition algorithms, image matching and camera tracking, spatial sound processing, speech communication by humans and machine, autonomous robot learning of foundational representations, developmental algorithms, cognitive architectures, markov decision processes, etc.

DIRAC joined forces with the IST-project CoSy (Cognitive Systems for Cognitive Assistants, http://www.cognitivesystems.org) to organise this five-day summer workshop. Both projects have the common goal of investigating cognitive processes of learning and understanding environments using data as retrieved by different sensor inputs. This first DIRAC summer workshop is acknowledged as a PASCAL related event by the PASCAL Network of Excellence (http://www.pascal-network.org/).

2. Organizational Efforts

A lot of organizational work has been carried out to establish the first DIRAC summer workshop since the second edition of the Icebreaker Workshop. Lecturers were contacted and engaged by both the DIRAC and CoSy project and a venue place was selected. KUL prepared a brochure about the summer workshop containing information with respect to deadlines, venue place, program, courses and lecturers. ETH has offered and finalized the set up of an online payment system for people registering to the summer workshop. Additionally, a form was created to allow registrations by fax as well. IDIAP set up and maintained a new website (http://www.diracproject.org/workshop-2007) to advertise the summer workshop and to keep track of the online registrations.

Support from the PASCAL network was requested and obtained for an administrative assistant (from ETH Zurich) to be present at the workshop to organize the registration, coordinate the different sessions and support any questions from the workshop participants which may arise. Final organizational tasks such as the excursion to the town of Bern and the welcome reception were taken care of by KUL.

The summer workshop has been successfully advertised through the EuCognition network and DIRAC's liaison network. 47 people registered for the workshop, which is more than the foreseen number of 40 people.

The Training Board agreed on initial conditions on which to reimburse DIRAC people for their summer workshop costs. The registration cost will be reimbursed and the initial plan of reimbursing travel (plane or train) to a maximum amount of 200 Euros was accepted. The latter item has been opened up for discussion again in order to accommodate people coming from far away.

3. Tutorial Sessions

The summer workshop courses cover the main sensory modalities in cognitive sciences, how they interact and can be fused. The workshop program will span five days with two main tutorial sessions a day. An abstract of these sessions is provided below.

Speech Communication by Humans and by Machine (Hynek Hermansky)
Spectral analysis of sounds is one of undisputed elements of early auditory processing. Spectrograph, introduced to a general scientific public after the Second World War, was developed to emulate this elementary capability and had significant and lasting effect on our view of acoustic world and especially on speech engineering. However, the understanding of processing of sounds in biological systems advanced considerably since the day of the Spectrograph. The talk will discuss some speech processing techniques that are based on evolving understanding of the role of spectrally localized dynamic temporal cues in human auditory perception.

Visual Object Recognition (Bastian Leibe , Tinne Tuytelaars)
Visual object recognition research has made considerable progress in recent years, to an extent that computer vision algorithms are gradually becoming applicable to challenging real-world recognition tasks. Many of those advances have come from a better understanding of local features that can be robustly extracted and matched under the difficult conditions encountered in such settings, including viewpoint and illumination changes, clutter, and partial occlusion.

This first part of the recognition tutorial will therefore focus on local features and how they can be used for recognition, both for specific objects and for object categories. We will introduce the concepts behind state-of-the-art interest point detectors and local region descriptors and will discuss several concrete implementations. We will then describe spatial models that can be used for recognizing familiar objects. Generalizing from specific objects to entire visual object categories, we will show how those models can be extended to cover the variability in both appearance and spatial layout. Finally, we will demonstrate how those concepts are applied in state-of-the-art object detection systems and discuss ways how those systems can be extended to additional dimensions of variability, such as scale changes and image-plane rotations.

Image Matching and Camera Tracking (Tomas Pajdla)
Image matching and camera tracking is a useful tools for self localization, scene modeling and recognition. The state of the art paradigm for image matching and camera tracking combines feature detection and selection, robust statistics, optimization and algebraic geometry to find corresponding points in images and to recover motion of the camera in space. We will explain the paradigm and its main components.

First, the state of the art image matching based on affine covariant feature detectors and descriptors will be presented. We will build on the previous course "Visual Object Recognition Lecturers" by Bastian Leibe, Tinne Tuytelaars. Secondly, camera models and their estimation from minimal number of points will be explained. We will show principles of constructing appropriate models to minimize the number of points needed to estimate them. We will provide an informal introduction to algebraic geometry necessary to understand basics of the problem. Finally, we will explain robust estimation techniques based on Random Sampling Consensus in general and its variations useful for camera tracking.

Spatial Sound Processing (Jörn Anemüller)
Spatial information of a sound field is captured by recording it with several receivers, such as the two ears of the human auditory system or the several microphones found in modern hearing aids. Subsequent processing permits to extract several parameters of the ambient acoustics, e.g., an estimate of the number of sound sources present, their positions relative to the listener, and even the direction where other speakers are facing. Signal enhancement techniques are also routinely based on this spatial information, enabling us to enhance desired signal components and suppressing interfering (noise) sources.

This tutorial will give an overview on the fundamentals and applications of perception and processing of spatial sound. We will outline the physics of sound field generation and the physiology and psychophysics of how our hearing system perceives spatial patterns. Technical approaches to analyzing and filtering spatial sound data will be outlined, including principles of microphone array beam-forming and recent approaches in the field of independent component analysis of sound signals.

Neural Mechanisms of Visual Object Recognition and Categorization (Rufin Vogels)
We will review the functional anatomy of the primate visual system emphasizing the ventral visual stream which is involved in the coding of object properties. Then we will discuss the responses of single neurons in the various ventral visual areas using a computational framework that distinguishes between the two essential problems of object recognition: invariance for image transformations (position, size, illumination and viewpoint) and selectivity for object properties. We will discuss experimental findings related to categorization of visual images and the effect of categorization learning on the representation in visual areas as well as in non-visual areas such as prefrontal cortex. Finally, we will discuss the coding of dynamic images of visual actions by single neurons in the prefrontal, parietal and visual cortex.

Autonomous Robot Learning of Foundational Representations (Benjamin Kuipers)
An intelligent agent experiences the world through low-level sensory and motor interfaces (the "pixel level"). However, in order to function intelligently, it must be able to describe its world in terms of higher-level concepts such as places, paths, objects, actions, other agents, their beliefs, goals, plans, and so on. How can these higher-level concepts that make up the foundation of commonsense knowledge be learned from unguided experience at the pixel level?

This question is important in practical terms: As robots are developed with increasingly complex sensory and motor systems, it becomes impractical for human engineers to implement their high-level concepts and define how those concepts are grounded in sensorimotor interaction. The same question is also important in theory: Does AI depend necessarily on human programming, or can the concepts at the foundation of intelligence be learned from unguided experience? This tutorial will describe recent progress on these questions, including the learning methods that support them.

Cognitive Architectures (Matthias Scheutz)
The goal of the tutorial is to give students a brief overview of past and ongoing research in cognitive architectures. The expected outcome is an appreciation of the utility of cognitive architectures for formulating theoretical principles underlying cognition and for building computational models and artificial cognitive systems using cognitive architectures (e.g., to test these principles of cognition by replicating and explaining human performance on cognitive tasks). The course will start by looking at the nature, role, and utility of building computational models of cognitive functions. It will then introduce the main cognitive architectures (including ACT-R, SOAR, and EPIC), while also briefly reviewing some non-symbolic architectures (e.g., like LEABRA).

Logical Representational and Computational Methods for Markov Decision Processes (Craig Boutilier)
Markov decision processes (MDPs) have become standard models for sequential decision problems involving uncertainty within the planning and probabilistic reasoning communities. This tutorial will provide a brief introduction to Markov decision processes and survey some of the recent advances that have been made in the concise and natural representation of MDPs using logical techniques; and computational methods that exploit this logical structure. Representations such as dynamic Bayesian networks, BDDs, and the stochastic situation calculus will be discussed.

Developmental Algorithms (Frederic Kaplan)
Have you ever thrown sticks and stones in the water as a child, just to find out whether they would float or not? Or have you ever noticed how much fun babies can have by simply touching objects, sticking them into their mouths, or rattling them and discovering new noises? It is these embodied interactions, experiences and discoveries and not only the organization of our brain that together result in intelligence. During the past five years, we have been working on algorithms that make robots eager to investigate their surroundings. These robots explore their environment in search of new things to learn: they get bored with situations that are already familiar to them, and also avoid situations which are too difficult. In our experiments, we place the robots in a world that is rich in learning opportunities and then just watch how the robots develop by themselves. The results show relevant analogies with the ways in which young children discover their own bodies as well as the people and objects that are close to them.

These tutorials will be given by lecturers well-known in their field of expertise. A short biography of each of the lecturers can be found in appendix A.

4. Group Discussion Sessions

In accordance with suggestions made by the reviewers, sufficient time was allocated for daily group discussions to allow participants to exchange ideas on various topics. The CoSy project took care of the content of these group discussions.

Groups were created in such a way that each group consists of a mix of people with different research backgrounds and from the two projects, DIRAC and CoSy. A senior researcher was assigned to each group to guide the discussions if necessary.

During the first half of the evening sessions groups worked individually on one of the daily topics. After individual discussions, groups reunited to a plenary session to present their thoughts, and ask the speaker/lecturer to comment on the main points.
Individual group discussions were primarily focused on the following questions.

- What are the general take home messages?
- What are the promising ideas and future prospects?
- What are easy or already solved problems? What can be done with recent state-of-the-art methods?
- What are open problems which can be solved within a PhD (3-5 years) or in 10 years? Which problems are too hard to be solved in the near future?
- What can you take away for your own research? What should other DIRAC/CoSy people take away for their own research?

However, the group sessions did not stay limited to discussing the above questions. Other related topics were discussed as well because of group members who are particularly interested in the topic or carry out research in a closely related field.

The following subsections summarize the most important points of the plenary sessions.

Day 1

During the plenary discussion, we have focused on solved and open problems of the two fields. We started with the topic of speech communication followed by visual object recognition. Unfortunately, Hynek Hermansky could not stay for the evening session, but other experts from the field helped to answer topic related questions. In general agreement we have concluded that solved problems, i.e., available working systems, are only successful in a constrained environment. Examples include dictation systems, that need to be customized to a specific person for good recognition performance, or automatic phone answering systems which are based on a very limited set of vocabulary.

Among the workshop participants many have used or know that speech recognition is being used in their research project. However, only a few actually do research in this field. We have discussed that current research focuses on relieving systems of the above mentioned limitation on vocabularies and speakers, as well as, being able deal with unknown words, or distinguish among different languages.

In visual object recognition open problems that come up during the discussion are similar to the ones mentioned in the lecture. We have agreed, that some problems, such as face recognition, or constrained OCR, such as license plate recognition are more or less solved problems. There are several commercial products that already take advantage of these algorithms. We have also discussed some similarities and differences between speech and visual processing.

Day 2

After the first day's discussions, we decided to alter the plenary session to focus more on questions that come up during the group sessions. In this way we find the speaker more active, which resulted in a more efficient exchange of knowledge between the participants and the experts. We kept this structure for the other days of the workshop.

On the second day three groups discussed image matching and camera tracking, while two groups focused on spatial sound processing. Concerning the first topic, the following questions came up during the plenary session:

- What is the relation between methods from robot localization to the presented methods of narrow and wide baseline camera pose estimation? Why, e.g., a dynamical model is not used in this context?

  On one hand, it is often easier to not have to assume anything about physical dynamics, at least not at the beginning. On the other hand, often there is already some kind of simple dynamics present, e.g., search regions for matching candidates are constrained based on previous frames.

- Can we incorporate high level semantic knowledge into the task, e.g., ‘’windows are part of a house”?

  On matching level, these kind of information is typically not used, however, there are opinions in the community, that we should just plug all information in to the system. These high-level information is typically incorporated at application level.

- Is the MatLab tool used during the lecture open source?

  Partially yes, but mostly no. But many implementations can be found on the Internet.

- Would a robot using an auto-focus zoom camera work?

  Not easy, but yes. The minval problem has to be changed to a 6-value problem (instead of 5).

- Are the algorithms real-time?

  Some parts are, yes (e.g. MSER). Many local feature extraction tecniques are or can be implemented on GPUs. However generally, several parts of the algorithms are computationally expensive (e.g. bundle adjustment).

- Can interest point-based features combined together with information from feature-free areas, where only dense sampling can be used, e.g., the sky?

  Dense features are hard problem themselves. Combination could be possible by, e.g., going from sparse feature locations to dense areas. But in general, it is a big problem.

- Are there any alternative applications similar to BouJou?

  Production companies, such as ones in the film industry, probably have their own application, which there are not willing to share because of obvious reasons. Image modeler is an alternative, but everything has to be done manually. We could not find available alternatives for automatic tools to BouJou.

- Any ideas how Google is planning to do their possible 3D-reconstruction project?

  We do not know. We only know that they are hiring the experts in the field.

Concerning the spatial sound processing presentation, the following questions were discussed during the plenary session:

- Are there any “optical illusions” for sound?

  Yes, there is a full CD on it. One example is the Shepherd tone.

- What are the main goals of the field?

  There are many different ones, including scene analysis, object recognition, direction of attention, etc. A currently popular PhD topic is, e.g., blind source separation.

- What are the reasons that these fields are developing orthogonally? Can engineering approach nature in this sense?

  It is very difficult. But there are exceptions. One such example is the mp3 encoding, that manages to reduce content by a factor of 10.

- What is the big picture?

  Fusion is important. Both DIRAC and CoSy projects’ goal is to build a bridge between researchers working in different modalities. Important to understand what can be done and what not. In vision we see a tendency from “low-level vision” (features, matching, etc.) to higher level, such as recognition.

Day 3

On Monday evening, the discussion sessions started with the topic on logical representational and computational methods for markov decision processes, followed by neural mechanisms of visual object recognition and categorization. We have kept the question-answer structure as the day before.

- Can the structure change online? What if something unexpected occurs? What if a transition occurs which according to your model is not possible?

  If you produce a value function, it will tell you what to do in every state. But it does mean your model was wrong, so you have to resolve this. This was not addressed in the talk. It is in the realm of robust optimization or partially observable models. You can optimize with respect to unknown parameters such as the transition probabilities which may not be known. Then you can have a family of possible models, and optimize with respect to those. If you are uncertain about the model, if you make a transition, it gives you some information about what the transition probabilities are. But it is a hard problem to incorporate that information.

- How can you establish a link between robotics and AI?

  For example, for manipulation. Its a hard question to answer in the abstract. If there is a robotics problem that is not navigation and localization, such as manipulation, there are continuous models and hybrid models. Such problems are best for using multiple conflicting objectives.

- Is art more real than reality itself? It seems that some drawings of objects produce a higher response than the objects themselves?

  If you have real objects, some parts of the objects may be hidden. But in a drawing you may see more features of the object. When you have an isolated part you have a very good response, but when you put together more parts, the response may disappear. If you apply Tanaka's response technique, a cell could respond to part B, but not part B and C together. On the other hand, when people show something that exaggerate some features such as in cartoons, or if you transform faces into caricature of faces, the response of the cells increase.

- What should be the main lessons for computer vision from biological vision?

  It would be interesting to try to instantiate system architectures similar to biological vision, because it can give some results on whether it works like the neuroscientists think or not.

- Let’s consider the caretaker scenario. If you are a caretaker and you are helping a person wash their hands. In the first 20 seconds you wait, but in the next 30 seconds you want to help them. How do you do this in a MDP given that it is independent of time?

  You assume it is a partially observable process. You can have time in the state and still be Markovian. Example, prompt the patient to squirt the soap on their hands. Discrete time model. 1s change of state. Fully observable. You have a variable for how much time has past since they were last prompted. It changes every time step: 1s, 2s. This if fully Markovian. You can condition your policy on it. Or you can have another state variable which is partially observable, modelling awareness or responsive state of the person, and if they do not respond in 1s, it gives you a noisy signal about their responsiveness. If they continue to not respond, your belief about their responsiveness will drop.

- What happens if an action takes very long, and then the system is just stuck in the transition between states? It seems that this framework is not appropriate to deal with such time considerations.

  You can not do this with discrete time models. You have to go to continuous time models, semi-DPs, and consider the density over duration of the action.

- You showed one study about cells in the hippocampus responding to semantic category? Can the same cell fire to another object in another setting?

  The impression is that cells do not respond to a given individual; they respond to a set of individuals. So it is sparse but not so sparse that it responds to one individual.

- How do the 'what' pathway and the 'where' pathway interact with each other? In cases where there is a latency between the dorsal and ventral streams, does the brain wait for the slower system before it performs its actions, or does it act based on the currently available information?

  Anatomically there are many connection between dorsal and ventral stream. AIP, LIP, IT recordings have been done to see how whether they communicate with each other. Dorsal stream is much faster than ventral stream. (40-50ms) Even in AIP and LIP, in IT 80-90ms. So the interaction would be more from dorsal to ventral, but we do not know for sure yet. One question regarding 2D shape is that since LIP knows 2D shapes before IT computes them, whether it communicates it to the IT.

- If you have two pieces of information, and one is faster than the other, does the brain have to wait until all the relevant information has arrived to act?

  Sometimes you have to react before all the information has arrived. That is why dorsal streams are fast, for fast reactions.

- There are many slew of neurons in IT. How does the brain create objects from all those features?

  It is a bag of features. Through development feature selectivity is built up, so it selects features for the object you want. How it is put together: if you look at selectivity, you do not do many further computations. A simple classiffier can get category information, size information. If you look at where the IT connects to, you can have a pretty good representation. The brain cannot average across trials. But you have more than one nerve. Many cells are firing to the same stimuli, so it is a bit like averaging. Many groups are looking into this. There is redundancy, so you combine cells and average out the noise. We average across trials but also take into account the variance across trials. Would that be enough to get rid of variability across trials? Yes, it should be.

Day 5

On Wednesday we have discussed developmental algorithms and cognitive architectures. During the former discussion we have also addressed several topics on autonomous robot learning of foundational representations. In the following we have summarized the most important questions.

- Is there some reason for not including other motivations? How would you do that?

  Including them is easy if you have your reward function and the progress measurement ability. The problem of course is with the choice of parameters. The objective here is to see what can be done with this single motivation.

- Would it not help to have a secondary goal?

  Yes, most researchers do that. They show that having intrinsic motivation helps reaching the goal faster and better than basic reinforcement learning. I feel it is slightly sad that they do that. We do not use exploration to achieve some other goal if it is exploration itself that we find interesting.

- Is there some relation between your work and Ben Kuipers's work? and what are the differences?

  The common idea is the focus on the bottom-up approach. Ben describes the work as building the basic blocks, and once they seem to be solved, he builds on top of that to solve more abstract problems. I do not believe in doing it that way because systems do not develop/perform in isolation. It may work with robotic tasks using laser range finders in corridors, but it may not work in general tasks. If the intermediary structure is false, the other systems built on top of that may not be appropriate.

  I like his work on building body maps, which is agnostic of the surroundings. But I do not like the way he uses it. He considers a special case, but with an Aibo, for example, if we detect all the sensors and make it move in its environment, we may believe the result to be a body map. I think what he has discovered is a signature of a given activity. Basically, we should not create abstraction when we are not sure that it is the right one.

- How would you include other value systems such as say the need for survival? How would you include that in your system?

  I would not want to do it. I think it would make the system more complex and hence more difficult to study. The fundamental requirement or feature of the current paradigm is the ability to make mistakes, to try out action the results of which are unpredictable. This may make this unsuitable for most practical applications. It may be used in tasks where exploration is a primary objective, for example a field of economics called optimal experimentation.

- Why are there so many architectures? Why is every group coming with their own architectures?

  Not every group comes with their own architecture; they typically use one of the more popular ones. Part of the reason is also need-driven. Initially people wanted to investigate only the principles under simulations. Later when they applied them to physical systems, people had to make modifications to suit their needs.

- What do you think of the CoSy architecture?

  I do not know much about the cognitive aspects of the architecture.

- What has changed between the various architectures: the room or the chairs?

That is, has the basic implementation changed much or is it just the implementation that has changed?

The room and the chairs both change. People look at the 12 features in different fashions, and interpret them in different ways. Over time the problems have evolved there by changing the necessities of the architecture.

- Can we ever subtract the effects of the software engineering issues from the cognitive architectures?

Most people who work on cognitive architectures do not care much about the software engineering aspects, except for the ones that are being used commercially. Typically the problems evolve over time, and so do the needs of the problem. As people understand them better, the architectures and the software engineering aspects also evolve over time.

Most people who work on architecture would claim that they only care about the development of the theory. They do not care about the software engineering aspects of it because they do not have to worry about the implementation on a physical system. Natural language processing is an example. Nobody developing cognitive architectures worries about how long it takes to parse the human speech and react to it, but in Human-Computer Interaction (HCI) tasks, it is essential to function in real-time. This was the main reason that forced us to develop our own version of a cognitive architecture.

- How do you compare different cognitive architectures on the basis of some basic features? Are there such basic features? What about figuring out how good they are on solving problems in general?

People do claim that different architectures are good for different tasks. A combination of different architectures is also commonly used. Part of the problem is that the usefulness of the architecture depends on the domain of the task. So there is no easy answer to this question.

- How could a cognitive architecture help to develop consistent models of sensory inputs over time? How do they represent the fact that the world is changing?

In both architectures I presented, knowledge technically sticks around for ever, but there is generally some degradation of information over time. There is some sort of belief revision system that gets updated over time.

- How could the symbolic systems help?

It can augment the data from other sensory modes, such as with visual input. For example, when detecting shapes, the shape could have been modeled as a symbolic label which can be compared against the current input in the visual field.

- Can there be some instance of active learning, where a mobile robot with the cognitive architecture performs an action explicitly to obtain some information to update its knowledge?

Yes, it is possible to do so, but it would depend on your model.

5. Conclusion

The EU projects DIRAC and CoSy have organized a joint workshop with a large tutorial component and daily group discussions in order to spread and discuss scientific results in their field of study. This Summer Workshop on Multi-Sensory Modalities in Cognitive Science was held at the Studienzentrum Gerzensee in Switzerland from 25 till 29 August 2007.

All necessary preparations were carried out. This event was properly advertised on the DIRAC website as well as through DIRAC's liaison network and the EuCognition network, resulting in all foreseen places to be filled up.

Appendix A

This appendix contains a list of the lecturers of the first DIRAC summer workshop including a short biography.

Bastian Leibe
(http://www.vision.ee.ethz.ch/~bleibe/)

Bastian Leibe obtained a MS degree in computer science from Georgia Institute of Technology in 1999 and a Diplom degree in computer science from the University of Stuttgart in 2001. From 2001 to 2004, he pursued his doctoral studies at ETH Zurich under the supervision of Prof. Bernt Schiele. He received his PhD degree from ETH Zurich in 2004 with his dissertation on "Interleaved Object Categorization and Segmentation", for which he was awarded the ETH Medal. After a one-year post-doc at University of Darmstadt in 2005, he joined the BIWI computer vision group at ETH Zurich in 2006, where he currently holds a post-doc position.

Bastian's main research interests include object recognition, categorization, and detection; top-down segmentation; and lately also tracking. Over the years, he received several awards for his research work, including the DAGM Main Prize in 2004 and the CVPR Best Video Award in 2006. He serves as a program committee member for ICCV, ECCV, and CVPR and is routinely reviewing for IEEE Trans. PAMI, IJCV, and CVIU.

Benjamin Kuipers
(http://www.cs.utexas.edu/~kuipers/)

Benjamin Kuipers holds an endowed Professorship in Computer Sciences at the University of Texas at Austin. He investigates the representation of commonsense and expert knowledge, with particular emphasis on the effective use of incomplete knowledge. He received his B.A. from Swarthmore College, and his Ph.D. from MIT. He has held research or faculty appointments at MIT, Tufts University, and the University of Texas.

His research accomplishments include developing the TOUR model of spatial knowledge in the cognitive map, the QSIM algorithm for qualitative simulation, the Algernon system for knowledge representation, and the Spatial Semantic Hierarchy model of knowledge for robot exploration and mapping. He has served as Department Chairman, and is a Fellow of AAAI and IEEE.

Craig Boutilier
(http://www.cs.toronto.edu/~cebly/)

Craig Boutilier is a Professor and Chair of the Department of Computer Science at the University of Toronto. He received his Ph.D. in Computer Science from the University of Toronto in 1992, and worked as an Assistant and Associate Professor at the University of British Columbia from 1991 until his return to Toronto in 1999. Boutilier was a consulting professor at Stanford University from 1998-2000, and has served on the Technical Advisory Board of CombineNet, Inc. since 2001.

Boutilier's research interests have spanned a wide range of topics, from knowledge representation, belief revision, default reasoning, and philosophical.

Frederic Kaplan
(http://www.fkaplan.com)

Frederic Kaplan is a researcher at the Ecole Polytechnique Federale de Lausanne (EPFL) in Switzerland . He graduated as an engineer of the Ecole Nationale Supérieur des Télécommunications in Paris and received a PhD degree in Artificial Intelligence from the University Paris VI. Between 1997 and 2006, he worked at the Sony Computer Science Laboratory in Paris on the design of novel approaches to robot learning and on the emergence of cultural systems among machines.

He published two books and more than 50 articles in scientific journals, edited books and peer-reviewed proceedings in the fields of epigenetic robotics, complex systems, computational neurosciences, ethology and evolutionary linguistics.

Hynek Hermansky
(http://people.idiap.ch/hynek)

Hynek Hermansky works at the IDIAP Martigny, Switzerland, and is a Professor at the Ecole Polytechnique Federale de Lausanne, Switzerland . He has been working in speech processing for over 30 years, previously as a Research Fellow at the University of Tokyo, a Research Engineer at Panasonic Technologies in Santa Barbara, California, a Senior Member of Research Staff at US WEST Advanced Technologies, and a Professor and Director of the Center for Information Processing at OHSU Portland, Oregon.

He is a Fellow of IEEE for “Invention and development of perceptually-based speech processing methods”, a Member of the Editorial Board of Speech Communication and of Phonetica, holds 5 US patents and authored or co-authored over 130 papers in reviewed journals and conference proceedings. He holds Dr.Eng. Degree from the University of Tokyo, and Dipl. Ing. Degree from Brno University of Technology , Czech Republic. His main research interests are in acoustic processing for speech recognition.

Jörn Anemüller
(http://www.anemueller.de/)

Jörn Anemüller studied at the University of London, England, and the University of Oldenburg, Germany, where he obtained the M.Sc. in Information Processing and the Ph.D. in Physics, respectively. He did a post-doctorate at the Salk Institute for Biological Studies and University of California San Diego in the field of neurobiological data analysis and is presently leading the speech processing effort within the Medical Physics Section at the University of Oldenburg.

Matthias Scheutz
(http://www.nd.edu/~mscheutz/)

Matthias Scheutz received the M.Sc.E. degrees in formal logic and computer engineering from the University of Vienna and the Vienna University of Technology, respectively, in 1993, and the M.A. and Ph.D. of philosophy in philosophy at the University of Vienna, Austria, in 1989 and 1995 respectively. He also received the joint Ph.D. in cognitive science and computer science from Indiana University Bloomington in 1999. He is an assistant professor in the Department of Computer Science and Engineering at the University of Notre Dame and director of the Artificial Intelligence and Robotics Laboratory.

He has over 90 peer-reviewed publications in artificial intelligence, artificial life, agent-based computing, cognitive modeling, foundations of cognitive science, and robotics.

His current research interests include agent-based modeling, cognitive modeling, complex cognitive and affective robots for human-robot interaction, computational models of human language processing in mono- and bilinguals, cognitive architectures, distributed agent architectures, and interactions between affect and cognition.

Rufin Vogels
(http://www.kuleuven.be/cv/u0006512.htm)

Rufin Vogels (Lab Neuro- en Psychofysiologie, KULeuven Medical School)has worked on the coding of object properties by single neurons in the macaque inferior temporal cortex and the neural mechanisms of visual categorization. He employs electrophysiological and behavioral techniques in rhesus monkeys to answer his research questions and has collaborated in several functional imaging studies in monkeys and humans as well.

His current interests are neural mechanisms of perceptual learning and categorization, neural adaptation and coding of visual actions.

Tinne Tuytelaars
(http://homes.esat.kuleuven.be/~tuytelaa/)

Tinne Tuytelaars received the MS degree in electrotechnical engineering at the Katholieke Universiteit Leuven in 1996. Since then, she has been working as a researcher in the computer vision group VISICS at that same university, which led to the PhD degree in 2000, for her work on “Local Invariant Features for Registration and Recognition”. Currently, she is a postdoctoral researcher of the Fund for Scientific Research Flanders (FWO).

Her main research interests are object recognition, wide baseline matching, and database retrieval, all based upon the concept of local invariant features. She serves as a program committee member for several of the most important computer vision conferences worldwide, and has over forty peer-reviewed publications.

Tomas Pajdla
(http://cmp.felk.cvut.cz/~pajdla/)

Tomas Pajdla received the MSc and PhD degrees from the Czech Technical University in Prague. He coauthored works that introduced epipolar geometry of panoramic cameras, investigated the use of panoramic images for robot localization, contributed to studies of panoramic mosaics, and studied non-central cameras and generalized epipolar geometries. He participated on developing an automatic approach to wide baseline image stereo matching and reconstruction of 3D scenes from many images. He coauthored works awarded the best paper prize at OAGM98 and BMVC02 and co-supervised the CMP team that won the second place for their 3D location estimation from uncalibrated 2D images in the ICCV 2005 Vision Contes