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

DELIVERABLE NO: D7.1
First DIRAC Summer School on Cognitive Engineering

Date of deliverable: 30.06.2007
Actual submission date: 24.07.2007

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)	

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.)

D7.1 First DIRAC Summer School on Cognitive Engineering

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 will be 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 has 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 have recently been taken care of by KUL.

The summer workshop has been successfully advertised through the EuCognition network and DIRAC's liaison network. Currently, 47 people are registered, which is more than the foreseen number of 40 people. The registration was closed on 20 May 2007.

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. Program

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.

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.

Visual Object Recognition (Bastian Leibe , Tinne Tuytelaars, Bernt Schiele and Ales Leonardis)
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, Bernt Schiele and Ales Leonardis. 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.

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.

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.

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.

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.

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).

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 the appendix below. In accordance with suggestions made by the reviewers, sufficient time has been allocated for daily group discussions. It is agreed with Bernt Schiele from the CoSy project that they will take care of the content of these group discussions and provide a summary of these discussions. An updated version of this deliverable including a summary of these group discussions will be presented one month after the end of the summer workshop.

4. Conclusion

The EU projects DIRAC and CoSy will organize 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 will be held at the Studienzentrum Gerzensee in Switzerland from 25 till 29 August 2007.

All necessary preparations have been carried out. The summer workshop program has been finalized and the required logistics organised. 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.

5. Appendix

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

Ales Leonardis
(http://www.fri.uni-lj.si/en/personnel/29/oseba.html)

Ales Leonardis is a full professor and the head of the Visual Cognitive Systems Laboratory with the Faculty of Computer and Information Science, University of Ljubljana. He is also an adjunct professor in the Faculty of Computer Science, Graz University of Technology. From 1988 to 1991, he was a visiting researcher in the General Robotics and Active Sensory Perception Laboratory at the University of Pennsylvania. From 1995 to 1997, he was a postdoctoral associate at the PRIP, Vienna University of Technology. He was also a visiting researcher and a visiting professor at the Swiss Federal Institute of Technology ETH in Zurich and at the Technische Fakultat der Friedrich-Alexander-Universitat in Erlangen, respectively.

His research interests include robust and adaptive methods for computer vision, object and scene recognition, learning, and 3D object modeling. He is an author or coauthor of more than 130 papers published in journals and conferences and he coauthored the book Segmentation and Recovery of Superquadrics (Kluwer, 2000). He is an Editorial Board Member of Pattern Recognition and an Editor of the Springer Book Series Computational Imaging and Vision. He has served on the program committees of major computer vision and pattern recognition conferences. He was also a program cochair of the European Conference on Computer Vision, ECCV 2006. He has received several awards. In 2002, he coauthored a paper, “Multiple Eigenspaces,” which won the 29th Annual Pattern Recognition Society award. In 2004, he was awarded a prestigious national Award for scientific achievements. He is a fellow of the IAPR and a member of the IEEE and the IEEE Computer Society.

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.

Bernt Schiele
(http://www.mis.informatik.tu-darmstadt.de/schiele)

Bernt Schiele is Full Professor of Computer Science at Darmstadt University of Technology since April 2004. He studied computer science at the University of Karlsruhe, Germany. He worked on his master thesis in the field of robotics in Grenoble, France, where he also obtained the "diplome d'etudes approfondies d'informatique". In 1994 he worked in the field of multi-modal human-computer interfaces at Carnegie Mellon University, Pittsburgh, PA, USA in the group of Alex Waibel. In 1997 he obtained his PhD from INP Grenoble, France under the supervision of Prof. James L. Crowley in the field of computer vision. The title of his thesis was "Object Recognition using Multidimensional Receptive Field Histograms". Between 1997 and 2000 he was postdoctoral associate and Visiting Assistant Professor with the group of Prof. Alex Pentland at the Media Laboratory of the Massachusetts Institute of Technology, Cambridge, MA, USA. From 1999 until 2004 he was Assistant Professor at the Swiss Federal Institute of Technoly in Zurich (ETH Zurich).

His main research interests are in computer vision, perceptual computing, statistical learning methods, wearable computers, and integration of multi-modal sensor data. He is particularly interested in developing methods which work under real-world conditions.

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 Contest.