Issue N° 2 / July 2006

Editor : Céline Aymon, celine.aymon@idiap.ch

DIRAC
Detection and Identification of Rare Audiovisual Cues

Newsletter

Contents

News

2006-06-15
Best Video award at ICC VPR 2006

NIPS
Neural Information Processing Systems Conference and Workshops Vancouver and Whistler, BC, Canada, December 4-9, 2006

The Call for Workshops and the Call for Demonstrations are now available online at: http://www.nips.cc

The Submission deadlines are as follows:
Call for Workshops:
August 4, 2006
Call for Demonstrations:
September 24, 2006

Cover Story

DIRAC's Training Program

DIRAC, as a multidisciplinary project targeting intensive cooperation between cognitive and engineering sciences as well as tending to overcome boundaries among auditory and visual research, will bring up a new generation of researchers, trained both in cognitive and engineering disciplines.

Researchers in the project have a variety of background and combine expertise in physiology of mammalian auditory and visual cortex and in audio/visual recognition engineering.

As of today, few if any educational curricula address this range of disciplines and as consequence there is a shortage of researchers who can collaborate effectively in cognitive and engineering sciences.

As an integrated project, DIRAC has the responsibility to "Spread Excellence". Its training program will make a significant contribution by providing opportunities for young scientists to work within the framework of an integrated project.

The placement scheme
The main activity within the training program is the placement scheme. Students and researchers are funded to spend periods of between 3 to 6 months working in DIRAC's labs on DIRAC-related projects. Each trainee has a designated supervisor in her/his host lab and is required to produce a project description at the start of the placement and a report at the end.
The placement program is not allowed to pay salaries; however it will provide travel expenses and a living allowance of €1,250 per month.
The scheme operates at academic levels ranging from undergraduate internships through masters and doctoral level students to post-docs. There is no nationality restriction on trainees.

Applications are dealt by the DIRAC Training Panel, which is chaired by Dr. Kurt Cornelis of the Katholieke University of Leuven (K.U.L)

In addition to gaining scientific knowledge and skills, DIRAC trainees will benefit from immersion into a state-of-the-art research culture. They will get experience of the whole research process, from proposal to publication. They will meet and interact with some of the best people in the field. They will see how scientific curiosity leads to engineering solutions.

The Summer school scheme
The training programs will also supports Summer schools, providing funding for invited speakers and for trainees to attend.

For more information on the training program, please visit our website: http://www.diracproject.org/trainingprogram/

DIRAC's first trainee

As of July first, DIRAC training program will host its first trainee. Joseph Keshet from the Hebrew University of Jerusalem will spent 3 months at the hosting site IDIAP, under the supervision of Professor Hynek Hermansky. He will work on research in speech sound spotting, addressing issues of context in fluent speech towards reliable spotting of sub-word units in speech.

Getting words from their sub-word parts (phonemes) may turn essential for forming hypothesis about low-probability but important words based solely on acoustic input while excluding contribution of prior language information. This, in conjunction with a more conventional speech recognizer, could provide means for identifying parts of speech containing unpredictable out-of-vocabulary words.

DIRAC c/o IDIAP Research Institute, Simplon 4, P.O. Box 592, CH-1920 Martigny
info@diracproject.org - www.diracproject.org
www.diracproject.org

Detection and identification of Rare Audio visual Cues (DIRAC) is an Integrated project funded by the EC's 6th Framework Program, managed by IDIAP (CH).

DIRAC
Detection and Identification of Rare Audiovisual Cues

Issue N° 2 / July 2006

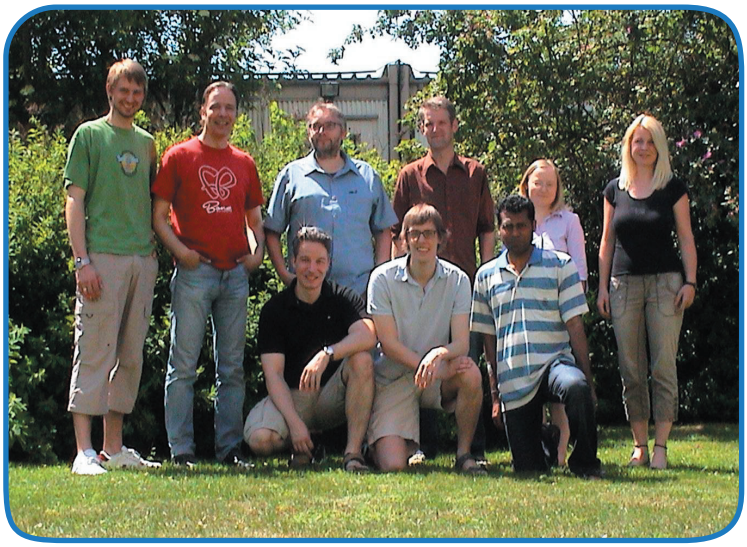
Editor : Céline Aymon, celine.aymon@idiap.ch

Newsletter

Leibniz Institute for Neurobiology, Magdeburg, Germany (LIN)

The Leibniz Institute ***(www.ifn-magdeburg.de)*** for Neurobiology in Magdeburg, Germany, is a member of Germany's Leibniz Association - a group of 84 research institutions, each one devoted to a particular research problem and typically bridging the traditional gap between fundamental and applied research in a certain field. Research at the Leibniz Institute for Neurobiology is devoted to illuminate neuronal mechanisms of learning and memory from the molecular level, via systemic levels, to the behavioural and psychological level. The particular focus is the study of adaptive processing of complex auditory stimuli such as speech.

Frank Ohl's group conducts research on neuronal activity in the auditory cortex associated with stimulus processing, learning-induced plasticity, and cognitive effort in rodents and humans. Our interests span from basic research on cortical neurodynamics to applied research which aims at the development of an interactive neuroprosthesis for the auditory cortex. Methodologically, the group focuses on electrophysiological recording of neuronal activities from single units, parallel recording of multiple single units, local field potentials and EEGs in awake behaving subjects. Our recent work has identified neuronal mechanisms that underlie the formation of concepts in cognitive auditory learning.



Members of the research group at the Leibniz Institute for Neurobiology, Magdeburg, Germany. From left to right: (front row) Lars Torben Boenke, Marcus Jeschke, Anton Ilango Michael, (back row) Max Happel, Achim Engelhorn, Frank W. Ohl (principal investigator), Matthias Deliano, Antje Fillbrandt, Kathrin Ohl.

Principally, our role for DIRAC is derived from the fact that animals' and humans' abilities to detect and identify rare audiovisual events and autonomously guide meaningful actions in response to them greatly surpass those of current non-living cognitive systems. The observability of especially animals' nervous system activities during such performances (a) allows the formulation of mechanistic models that can be used by engineers for implementing similar performance in artificial systems and (b) provides a unique opportunity to test whether suggested mechanisms derived by theoretical analysis are in fact implemented in living cognitive systems.

More specifically, the group's scientific efforts are rooted in work packages (WPs) 2, 4, and 5 as detailed below.

Our work for WP2 (auditory representations) focuses on the characterization of single units' spectrotemporal receptive fields (STRFs) in auditory cortex. This work has 2 facettes which are relevant to the general aims of DIRAC. First, the cortical STRF can be seen as a filter describing several important aspects of the transfer of sound stimuli into nervous system activity. While a few generic properties of STRFs have been clarified in the past many features of this descriptor, like its linearity, (spectro-temporal) separability and usability for predicting responses in general situations, are presently unclear. An even more severe restriction of current knowledge about this transform is that it is only derived for the steady state response under the assumption of stationarity. Cortical transfer functions, in contrast, appear especially sensitive to the occurrence of rare stimuli. The underlying mechanisms are presently not understood and are in the focus for our work for WP2. Second, the ultimate goal of all strategies of sensory processing is to guide meaningful actions of the system (organism or artificial) carrying these sensors. A key feature of such autonomous responses is the making of decisions between behavioural alternatives. Whereas classical views have seen this process as being separated from a «pure» sensory nature it is now clear that «sensory» processes, i.e. a system's responses to an incoming stimulus, are inherently entangled with the behavioural perspectives of that stimulus. We will use validated behavioural paradigms in combination with electrophysiological recordings of several relevant brain areas to build an integrated model of behavioural decision making in the context of rare, and in the usual sense «unclassifiable», stimuli. Our main cooperative partner in WP2 is OL.

Our work in WP4 (learning and categorization) focuses on a mechanistic understanding of learning-induced plasticity of STRFs and other observables of cortical neuronal activity. A particular focus will be on the neuronal mechanisms that lead to previously discovered sudden state changes of cortical activity that accompany the formation of new cognitive categories. Specifically, the impact of novel acoustic stimuli on the ongoing cortical dynamics will be studied; as this is the most likely candidate for driving the sensory system towards an instability point from which a new cognitive structure can be constructed that enables the organism to cope with an unexpected and/or rare event. Close cooperations with partner HUIJ are envisaged.

In WP5 (audiovisual information fusion) we address the question how organisms integrate information from multiple sensors to provide a meaningful representation of an inferred coherent source of the stimulus energy (the «perceptual object»). Our current work is based on our previous demonstrations of neuronal interaction between auditory and visual cortices in rodents during tasks in which events triggering one sensory modality (e.g. visual) have predictive value for upcoming events triggering another modality (e.g. auditory). A hypothesized strategy of how organisms deal with rare events is the inductive inference of potential meaning of the rare stimulus based on deductions performed on contextual cues, of which stimuli in another modality are an example. In our work we will specifically investigate whether and how cognitive concepts derived from experience in one modality are transferred to situations involving another modality and on which neuronal mechanisms this crossmodal transfer is based. The expected results both feed into and gain from a corresponding multi-stream model mainly developed by partners IDIAP and OHSU.

DIRAC
Detection and Identification of Rare Audiovisual Cues

Issue N° 2 / July 2006

Editor : Céline Aymon, celine.aymon@idiap.ch

Newsletter

Visual Representation in DIRAC project

Visual representations and the comparison of computational approaches with their counterpart in biological systems are the main focus of Work Package 3 (WP3). We put a first emphasis on the investigation of the visual perception of biological motion such as moving persons. Biological systems exhibit stunning capabilities for this task, however, it is not yet understood how the brain solves this problem. By the means of psycho-physical experiments (macaque), we study how the brain represents different categories of activities. These findings are of direct relevance for the design of computer algorithms that aim at solving similar visual recognition tasks. Both computational and psychophysical studies will share a common dataset with body motions and activities. A very powerful concept that is found in biological systems is the use of feedback from high-level semantic understanding to low-level visual analysis. Such cognitive feedback-loops can also render computer vision algorithms more robust and efficient. In this Work Package, a second emphasis is put on the identification and exploration of areas where cognitive loops can be built between different levels of visual processing. Of particular interest are combinations of `early' processing levels, with `higher' or semantic levels, like recognition.

DIRAC integration scheme
The algorithms we work on have many applications in modern life, e.g. in the form of devices with cognitive capabilities. We are particularly interested in devices that provide assistance for elderly people in indoor and outdoor scenarios, helping them to navigate and warn about unexpected and potentially dangerous events or situations. The combination of recognition, reconstruction and audio-analysis forms the basis for an integration of our activities in a prototype.

Biological motion
Researchers in neuroscience have proposed computational models of visual perception of biological motion in the brain. Such models are an interesting link between psychophysical studies of neural processes and the algorithmic approaches that are applicable in engineering. We try to establish correspondences between algorithms and such computational models of the brain. This is facilitated by the fact that we use the same kind of visual stimuli for neurophysical experiments and our algorithmic efforts.

On the computational side, we have implemented an initial version of a monocular tracker that operates on sequences of images of moving humans. The mapping from silhouettes to body configurations, which is of interest here, is often ambiguous, since there are multiple body poses that yield very similar silhouettes. A statistical model has been trained on a training database, consisting of pairs of silhouettes and corresponding body poses. This kind of model yields a probability distribution over possible body poses given an observed shape. In such a way, remaining uncertainties, e.g. when certain body parts can not well be estimated in a specific image, can be handled. So far, form/shape is the main cue from which the body pose is estimated, a temporal layer helps to improve the results.

On the psychophysical side we investigated whether macaque neurons represent the similarities among dynamic images of actions, images that combine form and motion information. For this, we employed a parameterized set of actions of human stick figures. The actions included 3 'prototypes' (lifting, knocking and throwing) and their blends. We used static presentations of snapshots and action movies to determine the contribution of form and motion information. Preliminary results show that neurons selective to dynamic action images are present. The large majority of the responsive neurons responded only to segments of the action sequence, but not during the whole action sequence. Most neurons responded selectively to the movement of the arm of the stick figure presented alone and thus did not require the whole body configuration. As a population, the neurons represented the physical similarities among actions, as shown by a non-linear dimensionality reduction analysis. We speculate that single neurons can code for motion trajectories and thus contribute to the visual analysis of novel dynamic images of actions. The population activity can represent the similarities among the different actions which is important for categorization of novel action stimuli. In this ongoing work, most of the motion was restricted to one arm. The next step we plan to use even more complex action stimuli in which all limbs are moving and to train the animals to categorize these movies. These novel images will be made in collaboration with ETH who will use the same motion sequences for their computational work.

Cognitive Loops
Computer vision finds itself at an exciting stage in its development. Gradually, the recognition of object classes, actions and events, and material types is becoming a reality. This creates the perspective of exploiting a pivotal principle in the architecture of the brain: feedback loops. Taking the DIRAC integration scenario as a case in point, 3D reconstruction and map building can become easier and more accurate when we know which kind of object is being reconstructed. In turn, recognition becomes easier and more reliable given a geometric scene context. In the first half year of DIRAC, a first example of such a cognitive feedback loop has been built in collaboration between ETH and KUL. It integrates object detection and 3D reconstruction in a 3D city modeling scenario and demonstrates those two areas' mutual benefit. In current work, this integration is being extended in several collaborations between CTZ, ETH, and KUL towards an application on omni directional cameras and for ultimate use in the DIRAC audio-visual integration scenario.

Best Video Award for DIRAC's work

KULeuven and ETHZurich received the Best Video Award for their joint work on cognitive loops which was presented at the International Conference on Computer Vision and Pattern Recognition, June 17-22 2006, New York, USA.

The awarded video demonstrates how a 3D city reconstruction module devised by the KULeuven is combined with a car recognition module devised by ETHZurich so that each module feeds information to the other module, allowing the combined system to overcome the weaknesses of each seperate module. The implementation of such a cognitive loop between modules led to a drastic increase in robustness, efficiency and accuracy of the resulting 3D city model. The video is accompanied by a technical paper:

Nico Cornelis, Bastian Leibe, Kurt Cornelis and Luc Van Gool: «3D City Modeling using Cognitive Loops»

DIRAC
Detection and Identification of Rare Audiovisual Cues

Issue N° 2 / July 2006

Editor : Céline Aymon, celine.aymon@idiap.ch

Newsletter

News & Events

IDIAP 15 Anniversary Workshop: Call for Presentation

IDIAP, as part of its 15th anniversary, will organize a workshop on September 12 and 13, 2006, to be held in Martigny Switzerland.

The institute invites talks from students and researchers addressing either theoretical or application issue in any of the following fields: Machine Learning, Speech Processing, Computer Vision, Information Retrieval.

The talks should be 15 minutes + 5 minutes questions. Deadline for abstract submission is 11:59pm (GMT) Tuesday July 11, 2006.

Please see *http://www.idiap.ch/ws15/* for more information on application, registration and possibility to receive a travel grant.

DIRAC's Publications

(http://www.diracproject.org/publications/)

Structure from Motion with Wide Circular Field of View Cameras

Branislav Micusik and Tomas Pajdla

IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 28, No.7, pp. 1135-1149, July 2006

Abstract: This paper presents a method for fully automatic and robust estimation of two-view geometry, auto-calibration, and 3D metric reconstruction from point correspondences in images taken by cameras with wide circular field of view. We focus on cameras which have more than 180° circ field of view and for which the standard perspective camera model is not sufficient, e.g. the cameras equipped with circular fish-eye lenses Nikon FC-E8 (183 degrees), Sigma 8mm-f4-EX (180 degrees) , or with curved conical mirrors. We assume a circular field of view and axially symmetric image projection to auto-calibrate the cameras. Many wide field of view cameras can still be modeled by the central projection followed by a non-linear image mapping. Examples are the above mentioned fish-eye lenses and properly assembled catadioptric cameras with conical mirrors. We show that epipolar geometry of these cameras can be estimated from a small number of correspondences by solving a polynomial eigenvalue problem. This allows to use efficient RANSAC robust estimation to find the image projection model, the epipolar geometry, and to select true point correspondences from tentative correspondences contaminated by mismatches. Real catadioptric cameras are often slightly non-central. We show that the proposed auto-calibration with approximate central models is usually good enough to get correct point correspondences which can be used with accurate non-central models in a bundle adjustment to obtain accurate 3D reconstruction. Non-central camera models are dealt with and results are shown for catadioptric cameras with parabolic and spherical mirrors.

3D City Modeling using Cognitive Loops

Nico Cornelis, Bastian Leibe, Kurt Cornelis and Luc Van Gool

International Symposium on 3D Data Processing, Visualization and Transmission, 3DPVT, June 14-16 2006, University of North Carolina, USA

Abstract: 3D city modeling using computer vision is very challenging. A typical city contains objects which are a nightmare for some vision algorithms, while other algorithms have been designed to identify exactly these parts but, in their turn, suffer from other weaknesses which limit their application. For instance, moving cars with metallic surfaces can degrade the results of a 3D city reconstruction algorithm which is primarily based on the assumption of a static scene with diffuse reflection properties. On the other hand, a specialized object recognition algorithm could be able to detect cars, but also yields too many false positives without the availability of additional scene knowledge. In this paper, the design of a cognitive loop which intertwines both aforementioned algorithms is demonstrated for 3D city modeling, proving that the whole can be much more than the simple sum of its parts. A cognitive loop is the mutual transfer of higher knowledge between algorithms, which enables the combination of algorithms to overcome the weaknesses of any single algorithm. We demonstrate the promise of this approach on a real-world city modeling task using video data recorded by a survey vehicle. Our results show that the cognitive combination of algorithms delivers convincing city models which improve upon the degree of realism that is possible from a purely reconstruction-based approach.

3D Reconstruction by Gluing Pair-wise Euclidean Reconstructions, or ``How to Achieve a Good Reconstruction from Bad Images"

Daniel Martinec, Tomas Pajdla,

International Symposium on 3D Data Processing, Visualization and Transmission, 3DPVT, June 14-16 2006, University of North Carolina, USA

Abstract: This paper presents a new technique for estimating a multi-view reconstruction given pair-wise Euclidean reconstructions up to rotations, translations and scales. The partial reconstructions are glued by the following three step procedure: (i) Camera rotations consistent with all reconstructions are estimated linearly. (ii) All the pair-wise reconstructions are modified according to the new rotations and refined by bundle adjustment while keeping the corresponding rotations same. (iii) The refined rotations are used to estimate both camera translations and 3D points using Second Order Cone Programming by minimizing the L-infinity-norm. The second asset of this paper is a criterion for evaluating importance of an epipolar geometry in influence on the overall 3D geometry. The measured correspondences are reweighted using the estimated importance in the bundle adjustment. The performance of the proposed method is demonstrated on difficult wide base-line image sets.

Fast Compact City Modeling for Navigation Pre-Visualization

Nico, Cornelis; Kurt Cornelis and Luc Van Gool

IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 17-22 2006, New York, USA

Abstract: Nowadays, GPS-based car navigation systems mainly use speech and aerial views of simplified road maps to guide drivers to their destination. However, drivers often experience difficulties in linking the simple 2D aerial map with the visual impression that they get from the real environment, which is inherently ground-level, based. Therefore, supplying realistically textured 3D city models at ground-level proves very useful for pre-visualizing an upcoming traffic situation. Because this pre-visualization can be rendered from the expected future viewpoints of the driver, the latter will more easily understand the required maneuver. 3D city models can be reconstructed from the imagery recorded by surveying vehicles. The vastness of image material gathered by these vehicles, however, puts extreme demands on vision algorithms to ensure their practical usability. Algorithms need to be as fast as possible and should result in compact, memory efficient 3D city models for future ease of distribution and visualization. For the considered application, these are not contradictory demands. Simplified geometry assumptions can speed up vision algorithms while automatically guaranteeing compact geometry models. We present a novel city modeling framework which builds upon this philosophy to create 3D content at high speed which could allow for pre-visualization of any conceivable traffic situation by car navigation modules.