Issue N° 9 / June 2008

DIRAC
Detection and Identification of Rare Audiovisual Cues

Newsletter

Contents

News

DIRAC Summerschool Volterra, Italy 15-19 September 2008

Cover Story

Incongruent Events

When General and Specific Classifiers Conflict

A key characteristic of intelligent cognitive systems is to respond appropriately in situations not previously encountered or perhaps unexpected in a particular situation. This ability to respond in a meaningful way to novel events is still much farther developed in living cognitive systems as compared to artificial cognitive systems. One of the reasons for this is that research on (both living and artificial) cognitive systems has traditionally focused on the mechanisms by which a system detects and exploits the typicality of a given situation. To advance beyond this point, a relevant step is the development of a suitable formalism that will enable a coherent description of novel event processing.

In pattern recognition we may encounter unexpected events for a variety of reasons. One example is the first appearance of an object from a new sub-class, as when a child, familiar with the concept of 'dog', sees a Pointer dog for the first time. Another example is the incongruent collection of parts, as illustrated in Fig. 1; most people react with pronounced surprise when they first see our 3-legged cat, or when seeing an elephant in a city street environment.

By definition, an unexpected event is one whose probability to confront the system is low, based on the data that has been observed previously. In line with this observation, much of the computational work on novelty detection focused on the probabilistic modeling of known classes, identifying outliers of these distributions as novel events. To advance beyond the detection of outliers, we observe that there are many different reasons why some stimuli could appear novel.

Let's consider distinct types of unexpected events, focusing on 'incongruent events' - when 'general level' and 'specific level' classifiers give conflicting predictions. In other words, we consider events whose probability when computed using different levels in an inherent hierarchy of representations is inconsistent.

Figure 1: Examples of unexpected arrangements of parts.

Unified approach

The set of labels represents the knowledge base about stimuli, which is either given (by a teacher in supervised learning settings) or learned (in unsupervised or semi-supervised settings). In cognitive systems such knowledge is hardly ever a set; often, in fact, labels are given (or can be thought of) as a hierarchy. We identify two types of hierarchies:

Part membership, as in biological taxonomy or speech. For example, eyes, ears, and nose combine to form a head; head, legs and tail combine to form a dog.

Class membership, as in human categorization – where objects can be classified at different levels of generality, from sub-ordinate categories (most specific level), to

to be continued on page 2

DIRAC c/o IDIAP Research Institute, Centre du Parc, Rue Marconi 19
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).

Audio-Visual Incongruency detection

basic level (intermediate level), to super-ordinate categories (most general level). For example, a Beagle (sub-ordinate category) is also a dog (basic level category), and it is also an animal (super-ordinate category).

The two hierarchies defined above affect the constraints on the observed features in different ways. In the class-membership hierarchy, a parent class admits higher number of combinations of features than any of its children, i.e., the parent category is less constrained than its children classes. In contrast, a parent node in the part-membership hierarchy imposes more severe constraints on the observed features than a child node. This distinction is illustrated on a simple "toy" example shown in Fig. 2. Roughly speaking, in the class-membership hierarchy (right panel), the parent node is the disjunction of the child categories. In the part-membership hierarchy (left panel), the parent category represents a conjunction of the children categories. This difference in the effect of constraints between the two representations is, of course, reflected in the dependency of the posterior probability on the class, conditioned on the observations.

In order to treat different hierarchical representations uniformly we invoke the notion of partial order. Intuitively speaking, different levels in each hierarchy are related by a partial order: the more specific concept, which corresponds to a smaller set of events or objects in the world, is always smaller than the more general concept, which corresponds to a larger set of events or objects.

Definition of incongruent events

Observation X (where X denotes the data) is incongruent if discrepancy exists between two classifiers: either the classifier based on the more general descriptions (level) accepts the X while the direct classifier rejects it, or the direct classifier accepts X while the classifier based on the more specific descriptions (level) rejects it. In either case, the concept receives high probability at the more general level, but much lower probability when relying only on the more specific level. The direct classifier is based on a probabilistic model (of given class) derived from training data without using the partial order relations. In case of two examples we have seen before:

- In the part-membership hierarchy (left panel of Fig 2): while the probability of each part (head, legs, tail) is high (since the multiplication of those probabilities is high), the 'dog' classifier is rather uncertain about the existence of a dog in this data. How can this happen? Maybe the parts are configured in an unusual arrangement for a dog (as in a 3-legged cat), or maybe we are seeing a dog with a cat's tail (or an elephant in a city street).
- In the class-membership hierarchy (right panel of Fig 2): while the probability of each sub-class (Afghan, Beagle, Collie) is low (since the sum of these probabilities is low), the 'dog' classifier is certain about the existence of a dog in this data. How may such discrepancy arise? Maybe we are seeing a new type of dog that we haven't seen before - a Pointer. The dog model, if correctly capturing the notion of 'dogness', should be able to identify this new object, while models of previously seen dog types (Afghan, Beagle and Collie) correctly fail to recognize the new object.

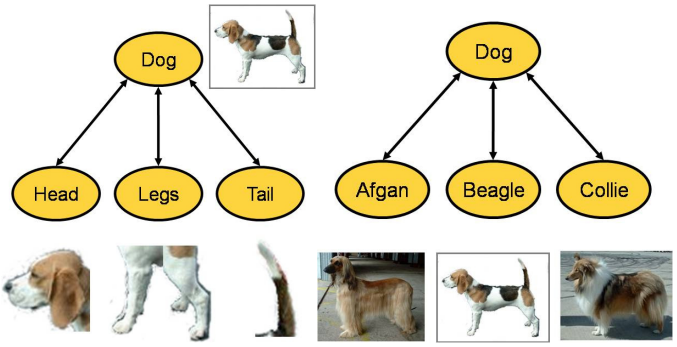

Figure 2: Examples. Left: part-membership hierarchy, the concept of a dog requires a conjunction of parts - a head, legs and tail. Right: class-membership hierarchy, the concept of a dog is defined as the disjunction of more specific concepts - Afghan, Beagle and Collie.

Neuromorphic Workshop'08, 20-29 April 2008

In April 20-29, Prof. Hynek Hermansky participated in the Neuromorphic Cognition Workshop in Italy. The program of the workshop and the participants can be found at ***http://www.ini.uzh.ch/~giacomo/sardinia08/***.

The main organizer of the workshop was the Institute of Neuroinformatics, ETH/UZH, Zurich. The workshop was the first in a series of yearly workshops that are intended to complement the long running Telluride summer workshop, mainly giving it more European flavor and also biasing it towards cognition.

Prof. Hermansky gave an invited tutorial presentation on main DIRAC theme and also chaired two sessions there.

Among many interesting contributions at the workshop, one of the most inspiring was a talk of Dr. Tony Bell (Redwood Center for Theoretical Neuroscience, Berkeley) on bio-inspired machine learning.

For the next year, it should be highlighted that the workshop provides support for students, selected from the applications, to be submitted sometime early in the calendar year.

The participation is recommended for those who might be interested in bio-inspired alternative approaches to machine learning. For more information, you may look at the intranet website ***http://www.diracproject.org/intranet/Talks_documents***

DIRAC c/o IDIAP Research Institute, Centre du Parc, Rue Marconi 19, P.O. Box 592, CH-1920 Martigny
info@diracproject.org - www.diracproject.org

DIRAC
Detection and Identification of Rare Audiovisual Cues

Issue N° 8 / June 2008

Newsletter

Audio-Visual Incongruency detection, first results

First results on audio-visual incongruency detection in DIRAC framework were obtained using AWEAR platform. Multimodal information streams present a related means to detect incongruous events within this framework. The unimodal classifiers are regarded as weakly constrained and their classification results are used as input for a «fusion» classifier. An incongruency between the unimodal streams are detected as the disagreement between the more constrained fusion classifier and one of the unimodal classifiers, provided that the unimodal outputs are obtained with a sufficiently high confidence score. In particular, a weighted linear combination of the hypotheses on different cues has been used. It has been shown in many cognitive and neurophysiology studies that humans use a similar approach for integrating multi-sensory inputs and integrate them in an optimal way. The incongruent events are first defined as different classifiers giving contradicting decisions, however, both with very high confidence. To interpret these incongruencies also requires some prior-knowledge, that is, to define a proper threshold so as to minimize the false alarm due to input noise, while maintaining a high detection rate.

First results for incongruency detection are obtained on data acquired at ETH last year :

- Data: 30 audio-visual speaker sequences (17 speakers, 7 male and 10 female) acquired using the AWEAR platform, see Figure 1.
- The speakers were asked to approach the camera and read a sentence about one minute long.
- The speech signals were captured by a head microphone worn by the actors. In a few sequences, the actors were asked to pretend an altered voice, that is, the male actors tried to speak with a high-pitched, female-like voice, and vice versa.

Figure 1: An example snapshot.

- We performed two kinds of experiments on the sequences, namely gender recognition and speaker verification. We found that integration of audio-visual cues could achieve better recognition performance than using a single modality alone, in particular under very noisy conditions. For example, in some of the sequences the illumination condition was very bad and the visual classifier gave many wrong decisions, thus provided low confidence in its output, while the audio classifier performed well and compensated for the weak classifier. The same effect was observed in the opposite direction when the audio chancel was noisy.
- In the gender recognition task, when the speakers were using altered voices, the audio gender classifier was usually «fooled» by the voice: It's output indicated high confidence for a wrong decision, while visual gender classifier gave the opposite decision again with high confidence. In the speaker verification task, we randomly selected 6 speakers as the trusted group, and the rest of the speakers belonged to the untrusted group. Our algorithms can accurately recognize all the speakers in the trust group. In addition, an ROC curve for unknown speaker verification was obtained by varying the detection threshold (see Figure 2). It shows that by integrating audio-visual cues we are able to achieve higher detection performance at lower false alarm rates.

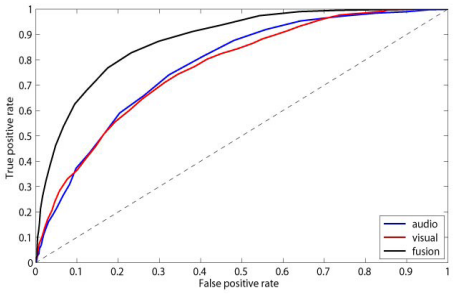

Figure 2: The detection threshold.

Cognitive Science Summerschool, Volterra, Italy, 15-19 September 2008

From 15-19 September, DIRAC will organize its second summer school in cognitive science. In cooperation with the POP (***http://perception.inrialpes.fr/POP/***) and COBOL (***http://www.cobol-project.eu***) projects, DIRAC has brought together a number of high level scientists (eg. Prof. Moshe Bar, Harvard Visual Neurocognition lab, USA, Prof. David Burr, Psychology department University of Firenze, Italy, Prof. Radu Horaud, Perception research group INRIA, France and Prof. Beatrice de Gelder, Department of psychonomics, Tilburg University, the Netherlands) from different disciplines and countries in order to provide sufficient food for thought for the students. Similar as last years summerschool the program intends to give students a chance to get together in an informal atmosphere and be able to discuss different topics, with a focus on (but not restricted to) multi-modality in cognitive science. Ample time will be given to ask questions to the lecturers and network with the other projects.

The setting of the summer school will be the SIAF conference center in the town of Volterra, an historical hilltop town with about 10,000 inhabitants, founded by the Etruscans in the heart of Tuscany three thousand years ago.

Deadline for early registration is July 1st and the registration fee is 600 euro, including 4 overnight stays, full board accommodation (breakfast, lunch, dinner and breaks), welcome reception, all courses and handling material.

For more information, you may look at the website ***http://www.diracproject.org/summer-school-2008/*** or contact Paul Konijn (***Paul.Konijn@esat.kuleuven.be***)

DIRAC
Detection and Identification of Rare Audiovisual Cues

Issue N° 8 / June 2008

Newsletter

Partner News

ACCV 2007 - Awards

Tobias Jaeggli, Esther Koller-Meier, and Luc Van Gool from ETH got awarded the "ACCV'07 Saburo Tsuji Outstanding Paper Award" for the following paper «Learning Generative Models for Monocular Body Pose Estimation».

The paper describes the human body pose estimation and tracking approach that ETH has developed for DIRAC and that will be used in collaboration with Rufin Vogel's neuroscience group at KU Leuven for joint experiments.

The tracking model is trained on motion-capture sequences recorded at ETH. The same sequences were used to create test stimuli for neurophysiological experiments with the macaque visual system at KU Leuven. Thus, both research groups are performing experiments based on the same initial data, and the successful body pose estimation results from our computational model can serve as verification that the stimuli set created at KU Leuven contains meaningful data.

At the same time, we hope that the neuro experiments at KU Leuven will result in hypotheses and predictions that can then be tested by the ETH computational models.

Telluride 2008

The 2008 Workshop and Summer School on Neuromorphic Engineering will be held from June 29th to Saturday, July 19th, 2008 in Telluride, Colorado.

Neuromorphic engineers design and fabricate artificial neural systems whose detailed architecture, design, and computational principles are based on those of biological nervous systems.

Over the past 12 years, this research community has focused on the understanding of low-level sensory processing and systems infrastructure; efforts are now expanding to apply this knowledge and infrastructure to addressing higher-level problems in perception, cognition, and learning.

The annual three week summer workshop will include background lectures on systems and cognitive neuroscience (in particular sensory processing, learning and memory, motor systems and attention), practical tutorials on analog VLSI design, mobile robots, hands-on projects, and special interest discussion groups.

For more information, you may look at the website ***http://ine-web.org/telluride-conference-2008***.

DIRAC's Publications

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

Journal papers

3D Urban Scene Modeling Integrating Recognition and Reconstruction
N. Cornelis and B. Leibe and K. Cornelis and L. Van Gool.
Supplying realistically textured 3D city models at ground level promises to be useful for pre-visualizing upcoming traffic situations in car navigation systems. Because this pre-visualization can be rendered from the expected future viewpoints of the driver, the required maneuver will be more easily understandable. 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. In this paper, we present a novel city modeling framework which builds upon this philosophy to create 3D content at high speed. Objects in the environment, such as cars and pedestrians, may however disturb the reconstruction, as they violate the simplified geometry assumptions, leading to visually unpleasant artifacts and degrading the visual realism of the resulting 3D city model. Unfortunately, such objects are prevalent in urban scenes. We therefore extend the reconstruction framework by integrating it with an object recognition module that automatically detects cars in the input video streams and localizes them in 3D. The two components of our system are tightly integrated and benefit from each other's continuous input. 3D reconstruction delivers geometric scene context, which greatly helps improve detection precision. The detected car locations, on the other hand, are used to instantiate virtual placeholder models which augment the visual realism of the reconstructed city model.

Conference papers

The Projectron: a Bounded Kernel-Based Perceptron
F. Orabona,J. Keshet,B. Caputo
to appear in Proceedings of the 25th International Conference on Machine Learning (ICMI), Helsinki, Finland, 2008

Hierarchical and Parallel Processing of Modulation Spectrum for ASR applications
Fabio Valente and Hynek Hermansky
ICASSP'08, Las Vegas, USA, April 2008

Exploiting Contextual Information for Improved Phoneme Recognition
Joel Pinto B. Yegnanarayana Hynek Hermansky and Mathew Magimai.-Doss
ICASSP'08, Las Vegas, USA, April 2008

Combination of strongly and weakly constrained recognizers for reliable detection of OOVs
L. Burget, P. Schwarz, P. Matejka, M. Hannemann, A. Rastrow, C. White, S. Khudanpur, H. Hermansky, J. Cernocky
ICASSP'08, Las Vegas, USA, April 2008

Confidence Estimation, OOV Detection and Language ID Using Phone-To-Word Transduction and Phone-Level alignments
C. White, G. Zweig, L. Burget, P. Schwarz, H. Hermansky
ICASSP'08, Las Vegas, USA, April 2008

Learning Generative Models for Monocular Body Pose Estimation
Tobias Jaeggli, Esther Koller-Meier, Luc Van Gool
in Asian Conference on Computer Vision (ACCV'07), Tokyo, Dec. 2007

A Mobile Vision System for Robust Multi-Person Tracking
Andreas Ess, Bastian Leibe, Konrad Schindler, Luc Van Gool
to appear in IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08), Anchorage, US, June 2008 (oral presentation)

Measuring camera translation by the dominant apical angle
Akihiko Torii, Michal Havlena, Tomáš Pajdla and Bastian Leibe
to appear in IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08), Anchorage, US, June 2008

Action Snippets: How Many Frames Does Human Action Recognition Require?
K. Schindler and L. van Gool
to appear in IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08), Anchorage, US, June 2008

Combining Densely Sampled Form and Motion for Human Action Recognition
K. Schindler and L. van Gool
to appear at DAGM'08 Annual Pattern Recognition Symposium, Muenchen, June 2008

Coupled Object Detection and Tracking from Static Cameras and Moving Vehicles
Bastian Leibe, Konrad Schindler, Nico Cornelis, Luc Van Gool
to appear in IEEE Transactions on Pattern Analysis and Machine Intelligence, accepted May 2008

DIRAC c/o IDIAP Research Institute, Centre du Parc, Rue Marconi 19, P.O. Box 592, CH-1920 Martigny
info@diracproject.org - www.diracproject.org