ANNUAL REPORT 2016

MAIN PARTNERS

CITY OF MARTIGNY

CANTON OF VALAIS

SWISS CONFEDERATION
The State Secretariat for Education, Research and Innovation (SERI)

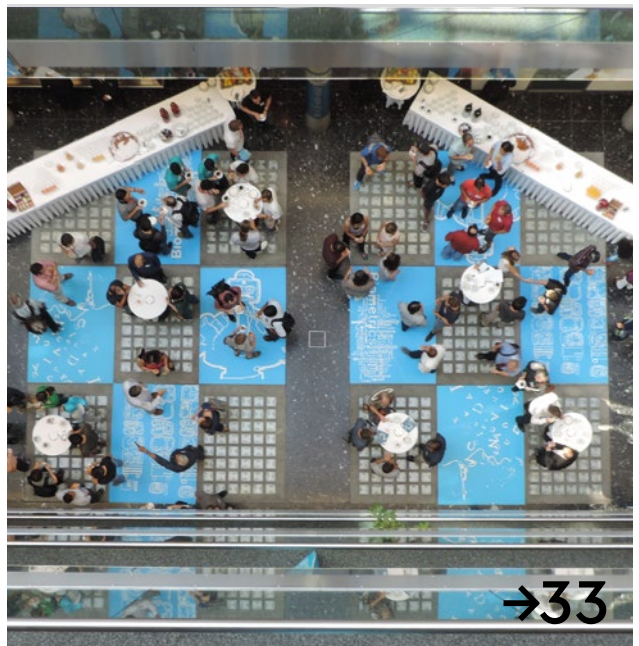|		
www.loterie.ch	www.swisscom.ch	www.groupemutuel.ch
www.epfl.ch	www.theark.ch	www.ideark.ch
www.snf.ch	www.kti.admin.ch	www.haslerstiftung.ch

CONTENTS

→24

→29

→31

→33

→25

→28

An anniversary— among successes and challenges

S imply, soberly, but not without a certain pride, Idiap celebrated its 25th anniversary in 2016— numerous events marking, throughout the year, this first quarter century (pages 32–33).

The institute now enjoys undeniable recognition in the field of artificial intelligence. This status is due in large part to the scientific skills and the qualities of organizer, manager, and visionary director embodied in Professor Hervé Bourlard. Motivating his teams, creative and demanding, he has managed—throughout his 20 years as this institute's director—how to combine these and other factors, with the result that today Idiap finds itself collaborating with the biggest names, including Google, Facebook, and Samsung. And all this, of course, without forgetting the nurturing of the academic network that links us to the four corners of the globe.

This magnificent success story is not, however, an end result in itself, but rather a platform on which to build in order to pursue our research, composed—as it is—of opportunities and innovations that meet society's needs. We have entered the era that some are referring to as the fourth industrial revolution. That era is broadly characterized by the digitization of knowledge and its wide diffusion, by the digitization of services, and by communication between objects—communication that takes place without direct human intervention.

The economic and philosophical effects of this new, hyper-technological world remain difficult to imagine. Some societal questions can, however, already be perceived. So, for example, everyone welcomes the training of scientists and technicians who can go on to design and create robots to perform tasks previously entrusted to a human operator. But, is it not also legitimate to engage in a debate with a view to guaranteeing a minimum income for those who find themselves without work? Or how should we view the innovations taking place in the healthcare sector—rich in promise, but fraught with dangers too? Yes, it is an extraordinary thing to be able to imagine living in great physical and mental shape until one is 100 years old or to have autonomous vehicles that take us to visit our doctor. But deep questions remain about our acceptance of such an "augmented man" in "a digitized society".

It is through this rapidly changing world that the staff of Idiap must find their collective way. Their dynamism and far-reaching skills will allow us to play our role and to overcome the inevitable setbacks. Idiap—a research institute and an innovative competence center in the field of computer science, especially with regard to interactions between people and machines—must nevertheless strengthen its collaborations and define its strategic position in the Swiss research and innovation landscape. It must also develop its contribution to the transfer of technology to the industrial sector, in collaboration with its privileged partner EPFL, and with the cantonal universities and universities of applied sciences.

And this is how Idiap will navigate, in serenity, toward its 50th anniversary.

"Idiap actively collaborates with the biggest names"

Olivier Dumas
President of the Foundation Council, Idiap

A look back at Idiap's origins

For some time now, "artificial Intelligence" has been hitting the headlines in both the daily and the generalist press, with regular references to the promise of the fourth industrial revolution. And behind this widely used expression lie the terms "deep learning" and "machine learning". As director of Idiap, I am delighted that the audacious wager placed 25 years ago has proven that the visionaries who founded Idiap, as well as the authorities that believed in them, were right.

When I arrived in Martigny in 1996, my goal was to create a truly multidisciplinary institute with an international standing, active in the fields of signal processing and machine learning—technologies common to the processing (recognition) of speech and vision by computers. At a time when multidisciplinary approaches were not yet an obvious choice, let alone a veritable obligation, I was given a free hand to pursue that goal, forging unique bonds between these important disciplines and integrating them—over the years—with other fields of research, such as multi-sensor analysis, multimedia information processing, and biometrics.

From fairly limited funding, and operating in the ultra-competitive world of science—where only an institution's reputation can attract the best researchers—developing a new research institute of the highest caliber has not been an easy task. Idiap's scientists have had to assert themselves and to fight to build our reputation and see our institute play in the big league. The confederation's decision to grant—in the shape of IM2—National Center for Competence in Research status to Idiap in 2001 contributed significantly to this success.

At its inception in 1991, Idiap was officially called the "Dalle Molle Institute for Perceptive Artificial

"Artificial intelligence: back to Idiap's roots"

Hervé Bourlard
Director, Idiap; Professor, EPFL

Intelligence" (Institut Dalle Molle d'Intelligence Artificielle Perceptive, IDIAP, in French)—today, I would rather refer to perceptual (or cognitive) artificial intelligence. Back then, the notion of artificial intelligence was at low point, a situation that favored the emergence of new methodologies of machine learning based on statistical models, notably the artificial neural networks that are so popular today (often grouped together under the name of "deep learning" or "deep neural networks"). And this is why the name "Idiap Research Institute" has gradually replaced the original acronym IDIAP. By a strange twist of fate, it is only now that this acronym is behind us and the five letters that make up the name "Idiap" stand alone in their own right that artificial intelligence has returned to the spotlight. I thank all of Idiap's staff for their active contribution to this veritable "comeback" and hope that we can keep building upon our achievements, fully exploiting our multidisciplinary culture.

organization

06 — 17

An institute on a human scale, but with an international reputation

With over one hundred employees, and research domains in touch with current challenges, Idiap is committed to scientific progress in the service of humankind.

In 1991, at its foundation, the Idiap Research Institute first established its mission of promoting scientific progress in the service of the well-being of humanity. Today, twenty-five years later, society's best interests remain at the heart of the institute's activities.

A profusion of technological products

The beginning of the 21st century has witnessed the arrival of a profusion of new technological tools. On the one hand, these tools have made possible considerable improvements in efficiency and comfort. On the other, they disrupt people's habits, leaving some users impoverished, and others weary of constant system modifications. It is within this context that Idiap works, essentially to improve man–machine interactions, and to optimize human communication.

A national and international network

In Switzerland, Idiap works mostly with the two federal institutes of technology, and with other universities of applied sciences and arts and cantonal universities at which research also takes place. Idiap is currently working on several European projects, and maintains close links with its numerous partners, including those in France, England, and Germany. Across the Atlantic, and in California in particular, privileged partnerships exist with the International Computer Science Institute (ICSI) in Berkeley and the Information Sciences Institute (ISI, USC Viterbi).

9

Structure

- Nonprofit foundation
- Independent institution
  (but connected to EPFL through
  a joint development plan)
- Specializes in the management
  of multimedia information and
  man–machine, multimodal interactions
- Founded in 1991 by:
  The city of Martigny
  The State of Valais
  EPFL
  The University of Geneva
  Swisscom

45% public
funds

5% provisions

50% research
projects
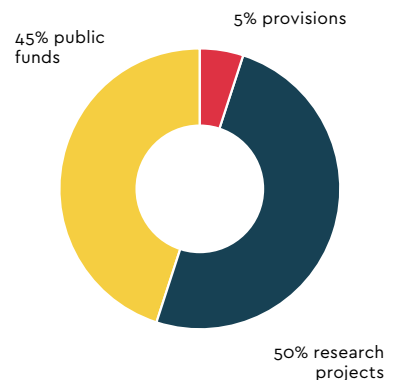
The three missions of Idiap

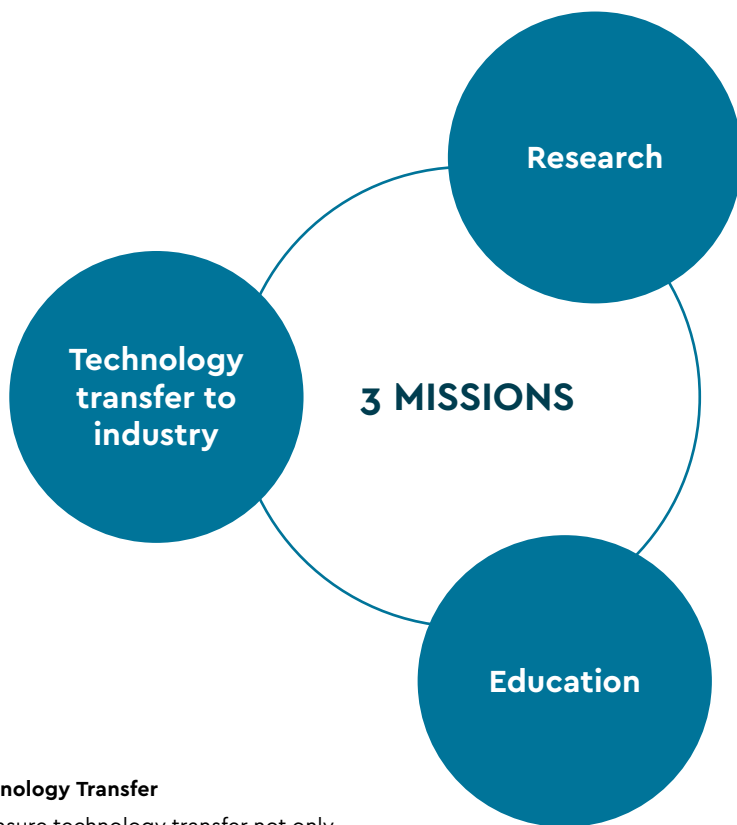The Idiap Research Institute is recognized by the Swiss confederation as part of the strategic domain of the Swiss federal institutes of technology and is supported under the Federal Act on the Promotion of Research and Innovation (RIPA) (Article 15).

The institute's vision is to promote quality of life through scientific progress made in relation to human and media computing. Idiap's activities are oriented in accordance with three complementary missions:

10

Research

Research
To conduct fundamental research projects at the highest level in its chosen fields of expertise, thereby assuring its position among the best, nation-, Europe-, and worldwide. On the international stage, Idiap benefits from a broad network of partners and collaborates actively with, among others, public and private research centers, and large universities. **Competitive research projects guarantee around 50% of the institute's financing.**

Technology transfer to industry

3 MISSIONS

Education

Education
To support the educational process by helping its interns discover the world of research. Idiap welcomes talented young researchers working toward their PhDs, and provides them with numerous courses at EPFL. **One in two scientific contributors is a research assistant.**

Technology Transfer
To ensure technology transfer not only by promoting the widest possible dissemination of its research results into the scientific community, but also—and most importantly—by forging close ties with the world of industry. **Technology transfer generates around 10% of Idiap's budget. Thanks to the incubator "The Ark" (IdeArk), Idiap contributes to the creation of numerous successful start-ups.**

Idiap in figures (2016)

Human resources

2	professors
2	senior scientists (MER)
8	permanent senior researchers
6	research associates
18	postdocs
31	research assistants
11	development engineers
6	system engineers
7	administrative staff
11	trainees (average/year)

136 people in total
94 full-time equivalents

- Switzerland 40%
- France 7%
- Rest of Europe 14%
- America 10%
- India 10%
- Rest of Asia 19%
- Stateless, 4 robots

11
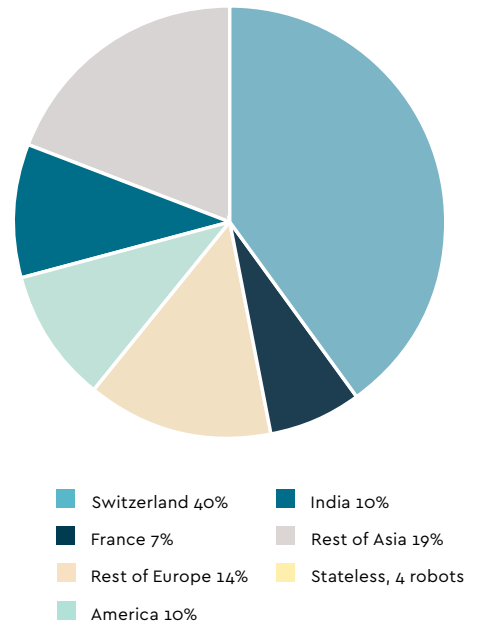
Scientific activities

- Participation in 68 research programs
- Project management in more than 48 consortia (Switzerland, Europe, industrial projects)
- Contribution to the economic development strategy of the Canton of Valais through the program The Ark and, in particular, the incubator IdeArk in Martigny
- More than fifty positions in IdeArk start-ups

Submission and funding of research projects

In 2016, Idiap submitted 58 projects (29 to Swiss funding agencies, 13 to European agencies, 15 industrial projects, and 1 US project). At the same time the institute was active in 68 projects.

Theses completed	6
Conference articles	76
Book chapters	7
Scientific articles	40

Publications

In 2016, Idiap researchers have contributed to 129 peer-reviewed publications.

58 submitted projects in 2016: EU 13, CH 29, IND 15, US 1

68 active projects in 2016: EU 11, CH 32, IND 24, US 1

5 research themes
10 application domains

Perceptual and cognitive systems	Social/human behavior	Information interfaces and presentation	Cyber-security	Machine learning
Automatic speech recognition, computer vision, handwriting recognition, multimedia document processing, robotics, machine translation.	Web and mobile communication, social interaction, social signal processing.	Multimedia information systems, person-alization and contextualization.	Person authen-tication; voice, face, and vein recognition; de-tection of crimi-nal activity.	Statistical modeling, neural networks, mathematical models.

Human–human, human–machine & human-robot interaction	Exploitation of rich multimedia archives	Collaborative and creative systems	Management of mobile systems	Energy management
Voice-controlled devices, voice–voice translation systems, navi-gation systems, gesture synthesis, learning from demonstration.	Semantic indexing, object recognition, audio-video content filtering, YouTube-document analysis, analysis of cultural heritage media.	Smart meeting rooms, video conferencing, multimedia index-ing and access, cross-lingual collaboration.	Signal processing for mobile plat-forms, mobile social networks, modeling of human activity to anticipate needs.	Heterogeneous energy networks, sensor networks, modeling of human activity to antici-pate needs.

Health and bio-engineering	Security and risk management	Cities of the future ("smart cities")	Entertainment and games	Home automation (domotics)
Smart patient (data-) management, protheses (auditory, corporal, etc.), bio-systems modeling, interfaces for the disabled, capture and in-dexing of medical documents, bio-imaging, microscopy.	Access control (for physical or vir-tual spaces), speaker identification, face identification, video surveillance, natural-risk mode-ling, industrial engineering, safety of public and private spaces.	Ecology, environmental management, pollution reduction, traffic reduction, better use of road networks, noise reduction.	Multi-lingual gaming, remote-family games ("togetherness").	Multi-sensor activity analysis, human-behavior analysis, adaptation to human behavior (e.g. heating), home safety and security.

Organization chart

Foundation Council

Scientific Advisory Board

Director
Hervé Bourlard

Deputy Director
François Foglia

Scientific College
F. Fleuret, D. Gatica-Perez, J.-M. Odobez

Admin. Committee

Research Committee

Director, Finance & Human Resources
Ed Gregg

System & Infrastructure
Frank Formaz

Speech and Audio Processing
Hervé Bourlard
Phil Garner
Mathew Magimai Doss
Petr Motlicek

Robot Learning and Interaction
Sylvain Calinon

Communication, Public Relations & Marketing
François Foglia

Development Engineers
Olivier Bornet

Computer Vision and Learning
François Fleuret

Uncertainty Quantification and Optimal Design
David Ginsbourger

Technology Transfer
Florent Monay
Hugues Salamin

Program Managers
Léonore Miauton

Social Computing
Daniel Gatica-Perez

Computational Bioimaging
Michael Liebling

Biometric Person Recognition
Sébastien Marcel

Perception and Activity Understanding
Jean-Marc Odobez

Natural Language Processing
Andrei Popescu-Belis

13

Research groups

Nine groups, each led by one or more top-level scientists, divide the institute's research activities between them. The group responsible for speech- and audio processing is led by the institute's director - Hervé Bourlard - and three permanent researchers. The remaining groups are all headed by permanent researchers including one professor titulaire and two adjunct faculties.

Speech & Audio Processing
Prof. Hervé Bourlard (second from right),
Dr. Phil Garner, Dr. Petr Motlicek, Dr. Mathew Magimai Doss
H-indexes respectively*: PG—22, PM—15, HB—57, MMD—23

Speech processing has been one of the mainstays of Idiap's research portfolio for many years. Today it is still the largest group in the institute, and Idiap continues to be recognized as a leader in the field. The expertise of the group encompasses statistical automatic speech recognition (based on hidden Markov models or hybrid systems exploiting connectionist approaches), text-to-speech, and generic audio processing (covering sound source localization, microphone arrays, speaker diarization, audio indexing, very-low-bit-rate speech coding, and perceptual background noise analysis for telecommunication systems).

Social Computing
Prof. Daniel Gatica-Perez
H-index: 50

Social computing is an interdisciplinary domain that integrates theories and models from mobile and ubiquitous computing, multimedia, machine learning, and social sciences in order to sense, analyze, and interpret human and social behavior in daily life, and to create devices and systems that support interaction and communication. Current lines of research include ubiquitous sensing of face-to-face interaction, behavioral analysis of social video, crowdsourcing, and urban data-mining using smartphones and mobile social networks.

Computer Vision & Learning
Dr. François Fleuret
H-index: 26

The scientific objective of this group is the development of new statistical learning techniques, mainly for computer vision, with a particular interest in their computational properties. Application domains include image recognition and scene analysis, the tracking of people and biological structures, and the understanding of images in general.

14

Perception & Activity Understanding

Dr. Jean-Marc Odobez
H-index: 37

This group conducts research into human–human activity analysis using multimodal data. This entails the investigation of fundamental tasks such as the representation, detection, segmentation, and tracking of objects and people, the characterization of their state, and the modeling of sequential data and the interpretation of that data in the form of gestures, activities, behavior, or social relationships. These investigations take place through the design of principled algorithms that extend models from computer vision, statistical learning, or multimodal signal processing. Surveillance, traffic analysis, analysis of behavior, human–robot interfaces, and multimedia content analysis are the main application domains.

Robot Learning & Interaction

Dr. Sylvain Calinon
H-index: 33

The Robot Learning & Interaction group focuses on human-centric robot applications. The scientific objective is to develop probabilistic approaches for encoding movements and behaviors in robots evolving in unconstrained environments. In these applications, the models serve several purposes (recognition, prediction, online synthesis), and are shared by different learning strategies (imitation, emulation, incremental refinement, or exploration). The aim is to facilitate the transfer of skills from end users to robots, or between robots, by exploiting multimodal sensory information and by developing intuitive teaching interfaces.

Uncertainty Quantification and Optimal Design

Dr. David Ginsbourger
H-index: 18

The Uncertainty Quantification and Optimal Design group focuses on quantifying and reducing uncertainties in the context of hi-fidelity models, with a main expertise in Gaussian process methods and the sequential design of computer experiments for optimization, inversion, and related problems. Application domains notably include energy and geosciences, with current collaborations ranging from safety engineering to hydrology and climate sciences.

Computational Bioimaging Group

Dr. Michael Liebling
H-index: 19

This group focuses on research in computational imaging and the analysis of biomedical images. This includes developing algorithms for image deconvolution and super-resolution in optical microscopy, three-dimensional tomography reconstruction from projections, and—more generally—combining unusual sensing devices and approaches with computational methods to produce images ideally suited to the observation and quantification of complex and live biological systems.

Biometric Person Recognition

Dr. Sébastien Marcel
H-index: 38

Biometrics refers, in computer science, to the automatic recognition of individuals based on their behavioral and biological characteristics. The Biometric Person Recognition group investigates and develops novel image-processing and pattern-recognition algorithms for face recognition (2-D, 3-D, and near-infrared), speaker recognition, anti-spoofing (attack detection), and emerging biometric modes (EEG and veins). The group is geared toward reproducible research and technology transfer, using its own signal-processing and machine-learning toolbox.

Natural Language Processing

Dr. Andrei Popescu-Belis
H-index: 22

The Natural Language Processing group studies how the semantic and pragmatic analysis of texts can improve the execution of two important tasks—machine translation and information retrieval. In particular, the group studies how the recommendation of information from networked multimedia repositories can be improved by analyzing a network's structure and the information content of its nodes.

*The h-index attempts to quantify both the productivity and citation impact of the publications of a given researcher based on the number of citations that they have received in other publications (citation level). The higher the index number, the more the citation level is important.

Foundation Council

The Foundation Council is responsible for the economic and financial management of Idiap. It defines the institute's structure, appoints its director, and–more generally–ensures Idiap's development and defends its interests.

Mr. Olivier Dumas, President
Independent manager
and business consultant

Mr. Patrick Furrer
Vice-Chancellor Research
and Innovation, HES-SO

Mr. Jean-Daniel Antille, Vice-President
Manager of the Regional Office
for the Economic Development
of French-Speaking Valais

Mr. Jean-René Germanier
Oenologist, Cave Jean-René
Germanier

Prof. Karl Aberer, Vice-President
Vice-President for Information Systems,
École polytechnique fédérale de
Lausanne (EPFL)

Mr. Jordi Montserrat
Regional Manager Venturelab

Mr. Marc-André Berclaz
Operational Director of the EPFL
Valais-Wallis Cluster

Prof. Christian Pellegrini
Senior Member of the Foundation Council of Idiap; Honorary Professor in the Faculty of Sciences at the University of Geneva

Mr. Stefan Bumann
"Head of the Service
des Hautes Ecoles (SHE)"

Mr. Dominique Perruchoud
President of the Board of Cimark SA

Mr. Marc-Henri Favre
President of the city of Martigny

Mr. David Watrin
Head of the product unit
Security & Intelligence at Swisscom

16

Advisory Board

The Advisory Board is composed of members of the scientific community selected by Idiap's management for their exceptional skills and avant-garde vision. Although their role is strictly advisory, their support is frequently sought and proves to be invaluable when making decisions regarding research, training, and technology transfer.

Dr. Alex Acero
Senior Director at Apple, Cupertino, CA, USA

Prof. Klaus-Robert Müller
Professor of Computer Science, TU Berlin. Director, Bernstein Focus on Neurotechnology, Berlin, DE

Prof. Dr. Vanessa Evers
Full Professor of Human Media Interaction, University of Twente, NL

Dr. Prem Natarajan
Director of the Information Sciences Institute, affiliated with the Viterbi School of the University of Southern California, USA

Prof. Anil K. Jain
Distinguished Professor, Department of Computer Science & Engineering, Michigan State University, USA

Prof. Bernt Schiele
Max-Planck Institute, Director, MPI Informatics; Professor at Saarland University, Saarbrücken, DE

Prof. Johanna Moore
University of Edinburgh, Head of the School of Informatics; Director of the Human Communication Research Center, UK

Dr. Luciana Vaccaro
Rector of the HES-SO, University of Applied Sciences and Arts, Western Switzerland, CH

Employees

SCIENTISTS

First name, last name, position, country of origin, residence, year of arrival

Oya Aran, research associate, Turkey, 2009

Afsaneh Asaei, postdoc, Iran, 2008

Dario Azzimonti, research assistant, Italy, 2015

Sushil Bhattacharjee, research associate, Switzerland, Renens, 2016

Joan Isaac Biel Tres, postdoc, Spain, 2008

Hervé Bourlard, director, Belgium/Switzerland, Saxon, 1996

Sylvain Calinon, researcher, Switzerland, Martigny, 2014

Gulcan Can, research assistant, Turkey, 2013

Olivier Canévet, research assistant, France, 2012

Tatjana Chavdarova, research assistant, Macedonia, 2014

Ivana Chingovska, research assistant, Macedonia, 2015

Tiago De Freitas Pereira, research assistant, Brazil, 2014

Subhadeep Dey, research assistant, India, 2014

Pranay Dighe, research assistant, India, 2014

Marc Ferras Font, research associate, Spain, 2011

François Fleuret, senior researcher, Switzerland, Yverdon, 2007

Kenneth Alberto Funes Mora, postdoc, Costa Rica, 2011

Philip Garner, senior researcher, United Kingdom, 2007

Daniel Gatica-Perez, senior researcher, Mexico, 2002

David Ginsbourger, senior researcher, France, 2015

Ioannis Havoutis, postdoc, Greece, 2015

Weipeng He, research assistant, China, 2016

Alexandre Heili, postdoc, France, 2015

Guillaume Heusch, research associate, Switzerland, Lausanne, 2015

Ivan Himawan, postdoc, Australia, 2016

Pierre-Edouard Honnet, research assistant, France, 2012

Rui Hu, postdoc, China, 2013

David Imseng, postdoc, Switzerland, St. German, 2009

Christian Jaques, research assistant, Switzerland, Lausanne, 2016

Noémie Jaquier, research assistant, Switzerland, Vevey, 2016

Cijo Jose, research assistant, India, 2014

Alain Komaty, postdoc, France, 2016

Pavel Korshunov, postdoc, Estonia, 2015

Tipaluck Krityakierne, postdoc, Thailand, 2015

Ilja Kuzborskij, research assistant, Latvia, 2012

Alexandros Lazaridis, research associate, Greece, 2012

Nam Le, research assistant, Vietnam, 2015

Remi Lebret, research assistant, France, 2012

Joel Legrand, research assistant, France, 2012

Yann Lepoittevin, postdoc, France, 2016

Michael Liebling, senior researcher, Switzerland, St-Sulpice, 2015

Ngoc-Quang Luong, postdoc, Vietnam, 2014

Srikanth Madikeri, postdoc, India, 2013

Mathew Magimai Doss, researcher, India, 2007

Parvaz Mahdabi, postdoc, Iran, 2014

Sébastien Marcel, senior researcher, Switzerland, Martigny, 2000

Olivia Mariani, research assistant, Switzerland, Muraz, 2016

Angel Martinez Gonzalez, research assistant, Mexico, 2016

Lesly Miculicich, research assistant, Peru, 2016

Amir Mohammadi, research assistant, Iran, 2016

Petr Motlicek, researcher, Czech Republic, 2005

Hannah Muckenhirn, research assistant, France, 2015

Skanda Muralidhar, research assistant, India, 2014

James Newling, research assistant, United Kingdom, 2013

Laurent Nguyen, postdoc, Switzerland, St-Légier, 2011

Olegs Nikisins, postdoc, Lithuania, 2016

Jean-Marc Odobez, senior researcher, Switzerland, Clarens, 2001

Pedro Henrique Oliveira Pinheiro, research assistant, Brazil, 2012

Dimitri Palaz, research assistant, Switzerland, Martigny, 2011

Nikolaos Pappas, research assistant, Greece, 2012

Novi Patricia, research assistant, India, 2012

Xiaojiang Peng, postdoc, China, 2016

Trung Phan, research assistant, Vietnam, 2015

Emmanuel Pignat, research assistant, Switzerland, Muraz, 2015

Andrei Popescu-Belis, senior researcher, Switzerland, Lausanne, 2007

Xiao Pu, research assistant, China, 2014

André Rabello Dos Anjos, research associate, Brazil, 2010

Dhananjay Ram, research assistant, India, 2014

Ramya Rasipuram, postdoc, India, 2010

Marzieh Razavi, research assistant, Iran, 2013

Dairazalia Sanchez-Cortes, postdoc, Mexico, 2012

Darshan Santani, postdoc, India, 2012

Adrian Shajkofci, research assistant, Switzerland, Martigny, 2016

Ajay Srinivasamurthy, postdoc, India, 2016

Ajay Tanwani, research assistant, Pakistan, 2015

Sibo Tong, research assistant, China, 2016

Sandrine Tornay, research assistant, Switzerland, Vernayaz, 2016

Raphael Ullmann, research assistant, Switzerland, Lausanne, 2012

Michael Villamizar, postdoc, Spain, 2016

Jonas Steffen Walheim, research assistant, Germany, 2016

Yang Wang, postdoc, China, 2016

Yu Yu, research assistant, China, 2015

- **Development Engineers**

Philip Abbet, senior development engineer, Switzerland, Conthey, 2006

Olivier Bornet, head of development team, Switzerland, Pont-de-la-Morge, 2004

Milos Cernak, senior development engineer, Slovakia, 2011

Samuel Gaist, development engineer, Switzerland, Ardon, 2013

Salim Kayal, development engineer, Switzerland, Vevey, 2011

Vasil Khalidov, senior development engineer, Russia, 2014

Christine Marcel, development engineer, Switzerland, Martigny, 2007

Florent Monay, tech transfer officer, Switzerland, Monthey, 2008

Alexandre Nanchen, senior development engineer, Switzerland, Martigny, 2008

Hugues Salamin, tech transfer officer, Switzerland, Dorénaz, 2014

Flavio Tarsetti, senior development engineer, Switzerland, Martigny, 2008

ADMINISTRATIVE STAFF

Elisa Bovio, junior program manager and financial assistant, Switzerland, Martigny, 2015

Martina Fellay, program manager, Austria, 2012

François Foglia, deputy director, Switzerland, Saxon, 2006

Edward Gregg, chief financial officer, United States of America, 2004

Léonore Miauton, head of program managers, Switzerland, Chexbres, 2012

Sylvie Millius, secretary, Switzerland, Vétroz, 1996

Nadine Rousseau, secretary, Switzerland, Saxon, 1998

- **System engineers**

Norbert Crettol, senior system administrator, Switzerland, Martigny, 2002

Bastien Crettol, system administrator, Switzerland, Martigny, 2005

Cédric Dufour, system engineer, Switzerland, Aigle, 2007

Frank Formaz, system manager, Switzerland, Fully, 1998

Louis-Marie Plumel, senior system administrator, France, 2011

Vincent Spano, webmaster, Switzerland, Martigny-Combe, 2004

INTERNS
First name, last name, country of origin, home institution

Idiap interns generally spend between six and ten months at the research institute. Some are students at the Ecole Polytechnique Fédérale de Lausanne (EPFL) and spend time at Idiap as part of their degree studies. Others come on student exchange programs set up with European projects in which Idiap participates.

Samaneh Abbasi, Iran, EPF Lausanne

Christian Abbet, Switzerland, EPF Lausanne

Krysten Ansermoz, Switzerland, EPF Lausanne

Afroze Baqapuri, Pakistan, EPF Lausanne

Maxime Bourlard, Belgium, University of Neuchâtel

Agata Cernakova, Slovakia

Kevin Chan, United States of America, University of California (US)

LongTao Chen, China, Nanjing Univ. of Science and Technology (CN)

Fabien Crepon, Switzerland, EPF Lausanne

Jaden Diefenbaugh, United States of America, Oregon State University (US)

Teodors Eglitis, Lithuania, Riga Technical University (LT)

Christopher Finelli, Switzerland, EPF Lausanne

Nikol Guljelmovic, Netherland, Delft Uni. of Technology (NL)

Maryam Habibi, Iran, Idiap Research Institute

Gil Luyet, Switzerland, University of Fribourg

Alan Maitre, Switzerland, EPF Lausanne

Zohreh Mostaani, Iran, Ozyegin University (TR)

Terry Niederhauser, Switzerland, HES-SO

Amrutha Prasad, India, Visvesvaraya Technological University (IN)

Nilavan Sathasivam, Switzerland

Amrollah Seifoddini Banadkooki, Iran, ETHZ, Zurich

Dominique Tao, Switzerland, EPF Lausanne

Véronique Thurre, Switzerland

Kelly Tiraboschi, Italy, University of Lausanne

Bogdan Vlasenko, Ukraine, Otto-von-Guericke University (DE)

Jing Yang, China, ETHZ Zurich

VISITORS
First name, last name, country of origin, home institution

Visitors are researchers or industry representatives who only spend a few days or weeks at the institute, some to strengthen inter-institutional relationships and others to get an insight into the work carried out at Idiap.

Yuxiang Li, China, EPFL

Catharine Oertel, Germany, Bielefeld University (DE)

19

finances
18 — 21

Profit and loss statement (CHF)

INCOME	2015	2016	%
City of Martigny	700 000	700 000	6.5%
Canton of Valais	1 720 000	1 720 000	16.1%
Swiss Confederation	2 467 300	2 423 700	22.6%
Loterie Romande	100 000	87 000	0.8%
Swiss National Science Foundation projects	1 715 249	1 537 844	14.3%
Hasler Foundation projects	606 689	175 781	1.6%
European Commission projects	1 395 595	1 404 423	13.1%
The Ark projects	225 070	245 380	2.3%
CTI projects	388 170	638 467	6%
EPFL contribution	72 000	72 000	0.7%
Industrial financing	538 818	414 669	3.9%
Other income/extraordinary income	396 274	738 592	6.9%
Provisions	—	555 000	5.2%
TOTAL INCOME	**10 325 165**	**10 712 856**	**100.00%**

EXPENSES			
Personnel expenses	7 884 132	7 948 334	74.2%
Education and travel	371 774	427 372	4%
Third-party expenses	167 114	102 292	1%
Computer equipment and maintenance	185 680	253 410	2.4%
Administrative costs	148 148	166 113	1.6%
Promotion and communication	45 468	45 769	0.4%
Rent	791 284	709 229	6.6%
Depreciation	263 690	363 090	3.4%
Exceptional expenses	—	—	—
Provisions	455 000	685 000	6.4%
TOTAL EXPENSES	**10 312 290**	**10 700 609**	**100%**
OPERATING PROFIT/LOSS	**12 875**	**12 247**	**—**

Accounting analysis 2016

For the fifth year in a row, Idiap is finishing this financial period (2016) with annual accounts exceeding CHF 10 million.

Public subsidies represent less than 50 percent of total revenues, thus respecting the strategic goal set by the institute's direction.

For a number of years now Idiap has focused on developing its relations with industry on the cantonal, national, and the international level, and on the transfer of technology. And clearly these efforts are bearing fruit. Over 10 percent of the budget is dedicated to projects closely related to the economy (funding sources: TheArk, CTI, and industrial).

Staff costs (salaries plus social charges) make up about three-quarters of total expenses.

Idiap closes its 2016 accounts with a surplus of CHF 12,247.

Federal, cantonal, and municipal subsidies

(In thousands of Swiss francs)

YEAR	2014	2015	2016	2017*
Confederation	2550	2467	2424	2418
Canton	1720	1720	1720	1645
Municipality	700	700	700	700

*Budget

Distribution of funding sources

- 7% City of Martigny
- 16% Canton of Valais
- 22% Swiss Confederation
- 1% Loterie Romande
- 14% Swiss National Science Foundation projects
- 2% Hasler Foundation projects
- 13% European Commision projects
- 2% The Ark projects
- 6% CTI projects
- 1% EPFL contribution
- 4% Industrial financing
- 7% Other income/extraordinary income
- 5% Provisions

Distribution of costs

- 75% Personnel expenses
- 4% Education and travel
- 1% Third-party expenses
- 2% Computer equipment and maintenance
- 2% Administrative costs
- 1% Promotion and communication
- 7% Rent
- 2% Depreciation
- 6% Provisions

Balance sheet (CHF)

ASSETS	31.12.2015	31.12.2016
Cash	3 467 573.13	5 313 117.83
Accounts receivable	319 121.25	130 394.85
Accrued income and other	622 013.91	434 238.84
TOTAL CURRENT ASSETS	**4 408 708.29**	**5 877 751.52**
Equipment	434 031.64	590 125.14
Financial assets	10 000.00	10 000.00
TOTAL NON-CURRENT ASSETS	**444 031.64**	**600 125.14**
TOTAL ASSETS	**4 852 739.93**	**6 477 876.66**

LIABILITIES	31.12.2015	31.12.2016
Accounts payable	349 413.85	213 106.26
Accrued expenses	1 878 638.93	3 747 835.83
Provisions	1 305 000.00	1 085 000.00
TOTAL FOREIGN FUNDS	**3 533 052.78**	**5 045 942.09**
Share capital	40 000.00	40 000.00
Special reserve	1 100 000.00	1 200 000.00
Retained earnings	166 812.11	179 687.15
Net income	12 875.04	12 247.42
TOTAL OWN FUNDS	**1 319 687.15**	**1 431 934.57**
TOTAL LIABILITIES	**4 852 739.93**	**6 477 876.66**

research
22 — 29

Distance learning 3.0

Today's decreasingly linear career paths and new technologies mean that knowledge has to be upgraded more and more rapidly by those wishing to maintain their skills and competitiveness. Continuing distance education offers an interesting alternative for individuals who wish to pursue training goals but want to reconcile those goals with their professional and family lives.

Distance learning offers recognized, quality education while adapting itself to the needs of the individual, and all without obliging prospective students to add to their already packed agendas the obligatory travel commitments that a return to the classroom normally requires. This type of quality continuing education can only be achieved through a robust, user-friendly platform that maximizes interaction between students and between students and their professors.

With a presence of more than 20 years in this field, and currently counting 1,500 active students, UniDistance has teamed up with Idiap and its spin-off Klewel in the context of the project eLearning-Valais 3.0. The project aims to develop the first ever solution for the production and use of multimedia and interactive resources for training purposes, taking advantage of the complementary skills of the three partners:

· UniDistance: pedagogical skills at the university level,
· Klewel: the acquisition, indexing, and sharing of multimedia presentations, and
· Idiap: the analysis of multimedia data.

Moving beyond massive open online courses (MOOCs), eLearning-Valais 3.0 aims to integrate, within a single digital solution, a set of features that already exist, but currently only separately:

· simplified production of recordings,
· intelligent navigation,
· multilingual indexing of recordings,
· generation of keywords, thanks to which any given document or any particular section of the curriculum can be easily located,
· sharing course notes with peers, and
· suggesting relevant files.

In the long run, this "Swiss Army knife"-like approach to teaching and learning in the digital age will also provide us with a better understanding of the collaborative processes that play out during distance learning. At the regional level, eLearning-Valais 3.0 will enable the Canton of Valais to strengthen its position as a leader in the field of distance learning, while meeting the social expectations and needs of students and diverse professional environments.

Idiap holds its course

"Renewed Confidence"

As is the case for many of the fields driven by federal policy making, the world of Swiss research moves to the rhythm of four-year cycles, for which budgets are discussed in advance. So, while the next of these cycles concerns the period 2017–20, in 2015 and 2016 already, important negotiations have taken place concerning Idiap.

For the next quadrennial period, the budget allocated by the confederation via Article 15 (b) of the Federal Act on the Promotion of Research and Innovation (RIPA) (Article 15, para. 3, letter b) will remain unchanged. For Idiap, this means nearly CHF 10 million of funding over four years (2017–20).

While the budget allocated to research in general, and to institutions that fall within the scope of Article 15 in particular, has risen very slightly, the confederation has decided to prioritize the strengthening of existing technological competence centers (Article 15, para. 3, letter c—RIPA), the creation of new ones, and the consolidation of research infrastructures.

Thus, while many institutions are facing budget cuts, Idiap can be proud to be maintaining a certain financial status quo. For Hervé Bourlard, director of Idiap, "In a rather tense context for public finances, I see—in the maintenance of our federal budget envelope—a continued confidence in our institute. And that envelope should be sufficient in the medium term for us to provide significant support for the institute's development."

Continuing diversification

Thus, buoyed by the unfailing support of the Canton of Valais and the city of Martigny, Idiap can pursue its plans for diversification and work, as planned, toward filling two newly created posts for permanent researchers—including one in the field of energy informatics. This should allow for closer collaboration with the Center for Research in Energy and Municipalities (CREM) in Martigny, and enable Idiap to participate in the development of Valais' energy ecosystem, working alongside strong partners like EPFL-Valais-Wallis and the HES-SO.

Idiap and EPFL

The beginning of this new budgetary period is accompanied by the signing, in March 2017, of a renewed joint development plan with EPFL. And while this plan aims to respect the confederation's wish that Idiap and other research institutes affiliated to EPFL be integrated into the Swiss academic network, it is more ambitious still—committing the two partners to collaborate on the development, at the national level, of certain of their joint activities.

Biometrics at the service of cybersecurity

We live in an increasingly connected environment and cybersecurity issues are, thus, not only of interest to specialists alone. Required to manage an ever-longer list of passwords, most of us generally choose the easiest way out—one simple password for multiple uses, sometimes even written down somewhere so we don't forget it… thus playing directly into the hands of cybercriminals.

Biometrics offers simple solutions to these problems. In fact, each human being possesses numerous physical particularities that make him or her unique—fingerprints, iris, face shape, venous network, voice… And it is in precisely this field that the Idiap-based Swiss Center for Biometric Research and Testing pursues its activities. Under the supervision of Dr. Sébastien Marcel, the center currently combines the talents of four doctoral students, three post-doctoral students, and three scientists.

 "We develop, in particular, authentication and identification technologies," Marcel explains. "For the former, a biometric trait—a "signature"— is compared to a reference, which makes it possible to verify whether or not we are who we claim to be. The latter is a question of finding the right correlation between an anonymous "signature" and another, known and stored in a database."

While the necessary information processing techniques are developed at Idiap, external collaborations—with schools, institutes, and private companies—are essential to the development and integration of suitable sensors (e.g., fingerprinting or vein readers).

Another aspect of cybersecurity concerns attempted identity theft. With respect to biometrics in particular, it is easy to imagine posing successfully as someone else by using a photo or a voice recording of the target person made without his or her knowledge or found on the Internet. Useful data like this could also be gained by pirating a biometric data bank. So technologies must be developed that make it possible to distinguish an original signature (e.g., vocal) from a copy.

In 2016, Idiap and the Biometrics Center organized two very different events. In January, the institute hosted the "Biometrics working group"—a first for Valais and for Switzerland, the event was attended by 87 specialists from around the world who debated norms and the implementation of ISO standards for the field of biometrics.

In a rather different context, in October, a stand set up at the Valais Fair—which attracted more than 200,000 visitors this year—offered an opportunity to promote the institute's biometrics activities and also to record the biometric data of more than 500 volunteers; data that will be used in future research projects.

Baxter—the robot that dresses you

Helping someone with reduced mobility to dress is the goal of i-Dress—a project of Dr. Sylvain Calinon's research group. The project is based around the robot Baxter, which is also used to test algorithms from projects as diverse as the development of an underwater robot or the creation of an artificial hand.

28

While an injured athlete can logically expect that his or her mobility will improve rapidly, the situation is quite different for the elderly, for whom mobility levels can worsen relentlessly. A robot tasked with helping one to dress should, therefore, ideally be able to adapt to the physical evolution of its charge, so as to provide him or her with personalized assistance at all times. Moreover, for such a robot to be usable in everyday life it must be relatively flexible, following—for example—the movements of its charge without either rushing or holding the person back. Finally, caregivers must be able to control the "medical" part of the programming, which must be as intuitive as possible.

In the face of these numerous challenges, the Baxter robotic platform is proving particularly useful. It is extremely flexible and can be taught a movement by imitation—that is to say, by guiding it physically, without resorting to a programming language.

With the help of CHIST-ERA funding, i-Dress launched in 2016, bringing together three research teams. The Idiap team, led by Sylvain Calinon, deals with the algorithms required for learning by imitation. Researchers at the Polytechnic University of Catalonia are working to determine the position of the clothing or footwear that has to be "handled". These objects, relatively large, but above all very deformable and therefore never quite identical, are relatively complicated to perceive. Finally, a team from the University of the West of England in Bristol is working on the human side of the project; that is, how future patients wish to interact with the robot. Designed four years ago, originally as an industrial robot, Baxter is used in research, where it is an ideal tool for prototyping and the testing of algorithms. This, indeed, is how Calinon uses Baxter in the framework of a project aiming to develop a remotely operated submarine robot (DexROV), and of research into the control of a prosthetic hand (TACT-HAND).

Idiap across the Atlantic
Prestigious collaborations in the US

In terms of how it finances its research projects, Idiap does not limit itself to Swiss and European sources of funding. While the institute enjoys the support of the Swiss National Science Foundation, Switzerland's Commission for Technology and Innovation, and the European Union's H2020 research programs, it is often also a partner of choice for US-based research initiatives.

In 2016, the biometrics specialists of Sébastien Marcel's group were involved in two projects funded, respectively, by the Defense Advanced Research Projects Agency (DARPA) and the Intelligence Advanced Research Projects Activity (IARPA). These two agencies have a combined annual research budget of around six billion US dollars. Their aim is to finance high-risk, high payoff projects capable of leading to technological advances of the highest order. These include the development of the Internet, the miniaturization of GPS systems, and the intelligent personal assistant Siri. These technologies were all initially developed for military applications before being made available to a wider public.

Between payoff and risk
These two institutions' calls to tender are highly selective and extremely competitive, to the extent that the number of laboratories that are likely to submit successful tenders is relatively low. Idiap and its Swiss Center for Biometric Research and Testing have been invited to join this highly exclusive club—testimony to their international reputations and to the success of the wager made by Idiap to develop its activities in the field of biometrics.

While the levels of funding allocated to these major projects are substantial—around USD 20–30 million—total project funding is never guaranteed. Indeed, each project sets several research teams in competition with one another, each working on the same thematic, and often the least advanced are eliminated from the project; a philosophy very different from that at work at Swiss and European research funding bodies. A further unknown, this time topical, is the recent change of administration—new people in charge can easily decide to review the agenda of the previous administration and to reorient its strategic axes. To handle such uncertainties, Idiap maintains an approach to risk analysis that ensures that the possible termination of any given project in no way jeopardizes the institute itself.

Cybersecurity and biometrics distinguished
Work on DARPA's project has already begun. Entitled SAVI, for "Spotting Audio-Visual Inconsistencies", and funded under the MediFor "Media Forensics" program, the project aims to detect the manipulation of audiovisual sources for the purposes of propaganda and to combat disinformation, or "fake news". In practical terms, the aim is to determine whether it is possible to detect differences between what we see when we consult an audiovisual document and what we hear. While the technologies employed are not part of the biometrics field per se, they are part of the overall theme of cybersecurity. The project is led by SRI International and is carried out in partnership with the University of Amsterdam.

At the end of 2016, Sébastien Marcel's team also received confirmation of the participation of Idiap and its Biometrics Center in the IARPA program Odin, the aim of which is to combat presentation attacks, or spoofing, aimed at deceiving biometric security systems. Idiap won the tender and will collaborate on the project, which will allow two new researchers to be recruited.

29

International deep learning workshop

From 4 to 6 July 2016, Idiap welcomed 300 researchers and students to a workshop on the methods and tools employed in deep learning—a booming discipline that will significantly modify our daily lives.

In October 2015, the computer program AlphaGo beats the European *go* champion 5–0. A few months later, it wins again, 4–1, this time beating the world champion—a feat that many scholars considered impossible. Indeed, the game of *go* is infinitely more complex than chess: when Deep Blue beat Kasparov in 1997, computers could only overcome relatively limited *go* players.

Behind these achievements lies what specialists refer to as deep learning—a discipline born in the 1950s under the name of neural networks, and which, thanks to developments in computational power, is currently experiencing a genuine revival. Idiap, long active in this field, is at the origin of Torch, one of the main development tools for a technology that is currently being used and optimized by Facebook.

Returning from one of the most important conferences devoted to deep learning, François Fleuret, head of the Computer Vision & Learning group at Idiap, decides to organize a workshop for his students. Calling on his connections, he invites software (Facebook and Google) and hardware (NVIDIA and AMD) specialists, and Professor Yoshua Bengio of the University of Montreal (third from the left on the picture) comes on board too, offering conferences focusing on the historical and contextual dimensions of deep learning. Enthusiasm for the workshop quickly spreads beyond the intimate circle of Fleuret's students, and in the end 300 participants from around the globe meet up in Martigny.

"From 1990 to around 2010, neural networks had lost their popularity," Fleuret explains. "Artificial intelligence research focused almost exclusively on other techniques that were more accessible to mathematical analysis. But these networks make it possible to efficiently exploit modern computers, especially graphic processing units (GPUs), which operate hundreds of times faster than conventional processors. These technologies make it possible to train computers using the immense quantities of digital data available thanks to the Internet, and—as we speak—are leading us to the very doors of a technological revolution that will change our daily lives."

Today, automatic recognition is one of the major themes of deep learning. From the abstract data of an image file, meaning is extracted. "Nourished by data, the machine learns little by little to differentiate a cow from a horse," Fleuret continues. "It's a process similar to that which allowed AlphaGo to become a master: having soaked up thousands of games, it assimilated the available strategies, and then trained itself by playing against itself, before beating the best of us."

I'm Pepper, the humanoid robot

Researchers at Idiap have acquired the robotics platform "Pepper". Their goal? Develop a social robot capable of interacting with humans in a complex environment.

Those familiar with Idiap and with the field of robotics need, we think, no further introduction to Nao. But since only recently, the small, self-contained, programmable humanoid robot developed by Aldebaran Robotics—which was taken over by Softbank Robotics in 2015—has a big brother, "Pepper", who stands 140cm high, head and shoulders above Nao's 48cm. And there is another anatomical difference: while Nao moves around on two legs, Pepper's creators have abandoned the bipedal approach, preferring three wheels, which give Pepper improved stability and greater mobility. Finally, a tablet computer situated on Pepper's torso allows it to increase its ability to interact with people it meets. In terms of both appearance and functionality, Pepper is very different from a robot like Baxter (see article on page 26). While Baxter has its origins in industrial applications, Pepper is a social robot, there to interact with humans.

Social robotics

And it was with that interaction in mind that Pepper was, in the summer of 2016, integrated into the Perception and Activity Understanding research group of Dr. Jean-Marc Odobez. Pepper will serve as an experimental platform for the new European project MuMMER (MultiModal Mall Entertainment Robot), led by a consortium of seven institutions. The aim of this project is to develop a robot capable of interacting socially with the customers of a large shopping center. At once entertainer, advertising medium, and customer service representative, Pepper will have to be able to operate in a complex environment, using its senses, or—rather—its sensors. And it is Pepper's perception of people—localizing them, detecting when they speak and their non-verbal behaviors and social signals like gaze or gestures—that Odobez's group is responsible for. At present, the group's researchers are working on analyzing the capabilities (and limitations) of the new platform's sensors and on improving its ability to understand human attention.

network

30 — 37

Coming home

The Valais-Wallis Ambition project has set itself the aim of encouraging citizens of Valais to return to their native canton once their studies are complete. Among the many advantages of this idea is that it enables the canton's economic community to benefit from the investments made by the State in the fields of research and training.

"Opportunity for young Valaisans"

Without its own university, the Canton of Valais is facing a worrying brain drain, despite the fact that it contributes in funding terms to the training of several thousand students per year. Indeed, according to a report published in November 2016 by the Higher Education Service of the canton's Department of Training and Security, during the previous year Valais had devoted CHF 61.5 million to its 5,712 students enrolled at universities, federal institutes of technology, or universities of applied science outside the canton (and CHF 55 million to the 5,599 students educated in Valais). There is also the question of whether those students studying outside Valais will return; particularly those who wish to pursue their academic careers at a university of applied science or a research institute based in the canton.

In order to encourage these students to come back to Valais, Idiap launched—in 2016—the initiative Valais-Wallis Ambition. Its aim is to offer young people who have completed their compulsory and secondary education in Valais an opportunity to embark on their university education by studying for a doctorate or a post-doctorate qualification at the Martigny-based institute. Students are selected based on their scientific qualities and on their suitability with regard to Idiap's

own chosen fields of research. As François Foglia, Deputy Director of Idiap explains, "the aim of the initiative is not to finance long-term research projects, but rather to connect our ongoing and future research aims with those candidates who are keen to return to the canton. One can also imagine, within the framework of this program, submitting a new research project based around the skills of a candidate accepted at Idiap. Valais-Wallis Ambition finances, in the medium term (a buffer period), the doctoral or post-doctoral student's salary, pending the official launch of a project supported by competitive funds, such as those of the SNF, the Hasler Foundation, or the CTI, or European funds, for example."

The program has enjoyed immediate success. In 2016, three doctoral students have already been hired under the auspices of Valais-Wallis Ambition. Their portraits can be found in the "Faces" pages of this report.

Idiap is 25!

To celebrate its first quarter century, Idiap turned to its partners and to the public at large. On 1–3 September, via three distinct events, the institute presented its activities and its know-how to almost 700 people.

The festivities began on Thursday, 1 September with an "Innovation Day". Mainly aimed at academia and industry, the event attracted nearly 200 attendees, to whom each Idiap research group presented its various projects.

The presentations complete, for more than three hours Idiap's guests walked the corridors of the institute, where more than 30 demonstrations, and multiple posters depicting some of the research institute's historic turning points, had been installed, thus experiencing at first hand the full range of Idiap's scientific and technology transfer activities.

The day ended with a conference given by Silvino Schlickmann—head of cybercrime research at INTERPOL—on the subject of the very real emerging threats that originate in the virtual world.

On the following day—Friday, September 2—Idiap invited the eight members of its renewed Scientific Council on the occasion of their biennial meeting. These international experts were thus also able to immerse themselves in the very core of the institute's research activities, before moving on to their scheduled discussions.

Finally, on Saturday, 3 September, the public—mainly natives of Valais—was invited to pass between the doors of Idiap, thrown wide for the day. No fewer than 450 people, including numerous children and families, came along, eagerly exploring the worlds of artificial intelligence, robotics, language processing, and biometrics.

In addition to these exceptional events, the institute also inaugurated its "5 to 6" program, which takes place on the second Wednesday of each month, and where the public is invited to a presentation and open discussion around Idiap's diverse research activities. Given the success of these gatherings, "5 to 6" will continue into 2017.

34

From idea
to start-up

Launched by Idiap in 2012, the International Create Challenge (ICC) gives young entrepreneurs the opportunity to transform their ideas into commercial prototypes during three weeks of immersion at the institute. In 2016, for its fifth edition, this unique program was once again an emphatic success, with—in the end—a total of four innovative projects rewarded.

From 31 August to 21 September 2016, Idiap welcomed eight teams of budding entrepreneurs—scientists, designers, and marketing and communication managers—including participants from Switzerland, France, Spain, and Brazil. At the conclusion of ICC'2016, the jury—composed of leading figures from the worlds of research, venture capital, and innovation—rewarded four projects that combine innovation and high scientific quality: Viprodev, Veintree, MeasureMe, and Dental Simulator.

ICC'2016 participants

1. VIPRODEV

TEAM

- Dr. Kadir Akin—postdoc at the Microelectronic Systems Laboratory (LSM) of EPFL (Switzerland)
- Georgel Vilsan—Master of Public Sector Management (Romania)
- Dr. Pablo Garcia—PhD in Computer Systems Engineering at the Universidad Complutense de Madrid (Spain)
- Luciano Cerqueira Torres—PhD in Strategy, Program, and Project Management at the École Supérieure de Commerce de Lille (France)

GOAL

VIPRODEV aims to develop a high-performance, stereoscopic distance measuring device as well as a graphical interface for those who use it. Named Depthor, the device uses a combination of RGB and depth cameras for image processing applications.

Viprodev team

2. VEINTREE

TEAM

- Christophe Bron—Dr. med. MSc.; currently manager of VTIP Sàrl and president of Groupe Veintree (France)
- Mathieu Rossi—IT engineer and manager of R4M (France)
- Didier Launay—vice-president of XMP Entrepreneur (France)
- Luc Rodrigues de Magalhaes—developer at Damavan Imaging SA and engineer at UTT (France)

GOAL

Develop a biometric authentication system based on the venous network in order to guarantee patient follow-up in the context of humanitarian crises, where information on the identity of victims is incomplete or even non-existent.

3A. MEASUREME

TEAM

- Magali Leyvraz—postgraduate at the Institute of Social and Preventive Medicine at CHUV (Switzerland)
- Gustavo Rios—BSc and MSc in Mechanical Engineering at EPFL (Switzerland)
- Eric Nguyen Van—BSc and MSc in Mechanical Engineering at EPFL (Switzerland)

GOAL

To develop a mobile application to determine the weight of a child and its state of malnutrition by means of images, with the aim of facilitating the in situ work of caregivers, and of assembling databases on malnutrition, thus enabling the early detection of nutritional crises without data having to be entered manually.

3A. DENTAL SIMULATOR

TEAM

- Professor Leandro Pereira—dental surgeon at the São Leopoldo Mandic Dental School (Brazil)
- Rodrigo Dias Takase—software coordinator at the Educational Technology Laboratory of the University of Campinas (Brazil)
- Matheus Rocco—IT student at the University of Campinas (Brazil)

GOAL

Develop a mobile application that allows students of dentistry to simulate, in a virtual environment, interventions such as anesthesia, root work, and the introduction of implants.

Idiap meets tomorrow's engineers

Because the children of today are the engineers of tomorrow, Idiap is committed to presenting its work, and the various professions practiced at the institute, to young audiences.

"The festival Hérisson sous gazon"

For the festival Hérisson sous gazon, which took place in Charrat on 18 and 19 June 2016, Hugues Salamin and Florent Monay put on a workshop entitled "The School for Robots". Their goal? Show children aged 6 to 14 how robots learn and introduce them to the concepts of programming, machine learning, and artificial intelligence, thus presenting Idiap's activities and skills to this very special audience. In the course of the workshop, robots had to learn to recognize noises, and successive groups followed one after another throughout the day, to the sound of the children's imitations of the cries of seals and of other sounds.
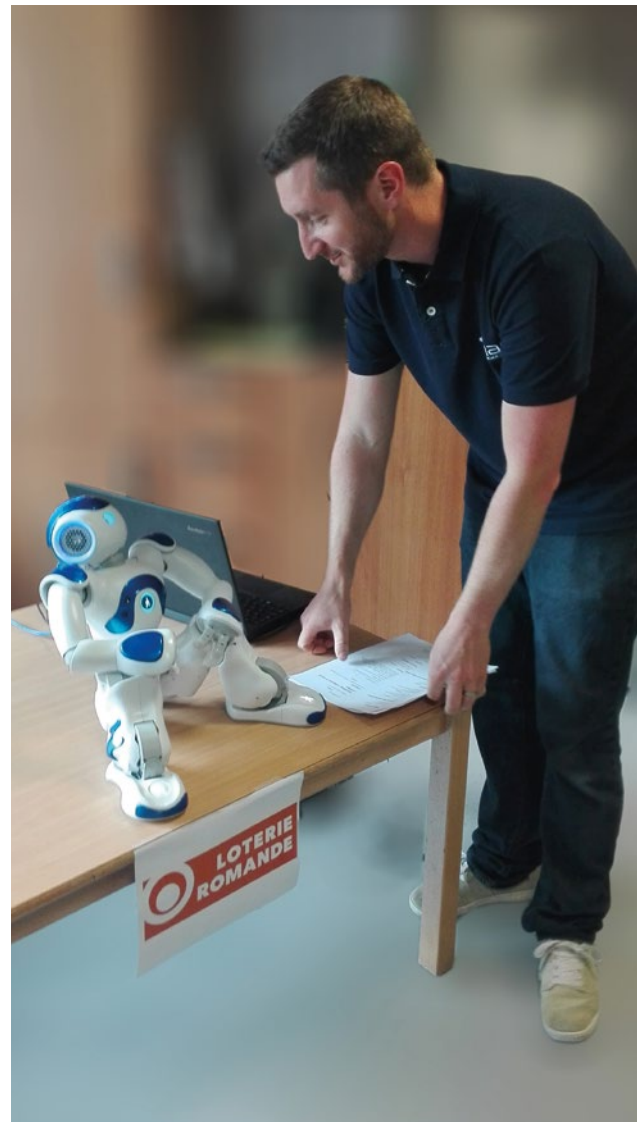
"Oser tous les métiers/Futur en tous genres"

The day-long event *Oser tous les métiers/Futur en tous genres* (Dare any Profession/All Kinds of Futures) is a father–daughter/mother–son exchange day that aims to counter the manner in which roles and professions appear to be compartmentalized with regard to gender. Its main objective is to enable girls and boys to explore new occupational fields and to broaden their horizons in preparation for the choices to come, so that girls do not limit themselves to "traditionally female" sectors and boys to "traditionally male" sectors—a very important goal given Idiap's "traditionally masculine" field of activity.

During the 10 November event, organized by the Valais Secretariat for Equality and the Family, Idiap welcomed 22 children, including three boys—all in their eighth year of studies—and explained the range of professions practiced at the institute: from researcher to system administrator, from development engineer to program manager. After a presentation of Idiap and the relevant professions, the children were able to roll up their sleeves and participate in four 90-minute workshops. On the program: the server room and the Idiap showroom, the robot Nao, Lego's WeDo, and—finally—a workshop on real-time image processing. All in all, a very positive experience that Idiap has been pursuing since 2012 and via which the institute hopes to inspire a new generation to take up vocations as future engineers.
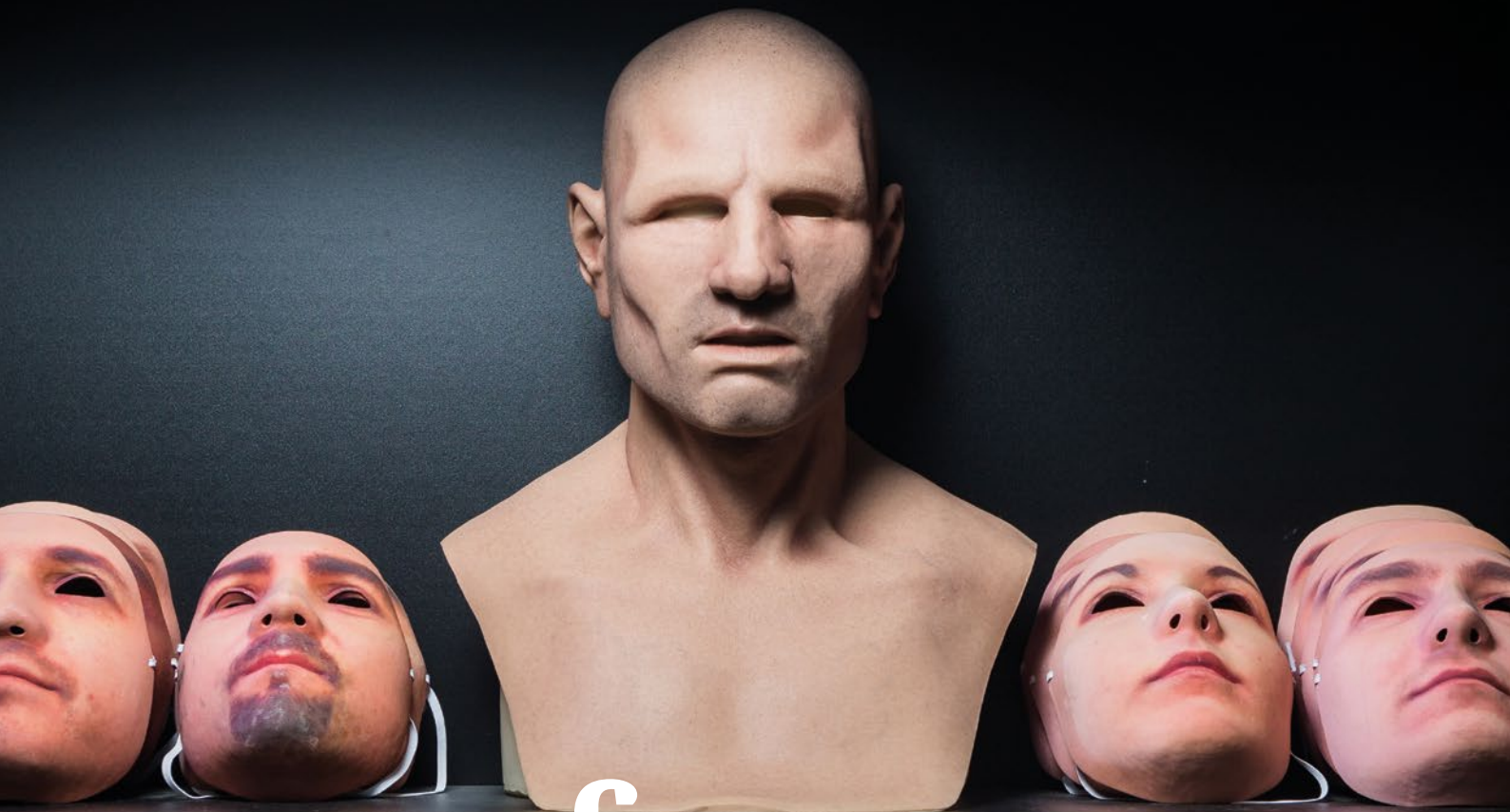
faces

38 — 43

Dribbling with the numbers

Having arrived at Idiap in 2004, Ed Gregg is now responsible for the institute's finances. But it was as a professional basketball player that the American first discovered Switzerland.

Ed Gregg owes his arrival in Europe more than a little to his considerable height—2.13 meters—but especially to his qualities as a talented basketball player. Indeed, it was Ed the athlete who embarked for the Old World back in 1976. He roamed the courts of a number of teams—playing twice in the European Cup—before coming to play in Switzerland, and then coaching several teams here.

Ed is Idiap's chief financial officer (CFO). Having studied Finance at Utah State University, Ed—now 63 years old—decided to return to the discipline when a new turn in his career path beckoned. So, while coaching in Ticino, he joined a fiduciary firm, before eventually meeting up with his son in Martigny twelve years later. In search of a job, Ed's path crossed that of Idiap. And when the institute embarked on a number of European projects, the fact that English is Ed's mother tongue convinced Pierre Dal Pont, former head of finance at the institute, to hire him. To begin with, Ed worked at

20 percent, then at 30, then 40, and so on, until he took over the position of CFO in 2010. While the financial reality of academia is, he finds, very different from that of the private sector, he feels at home in the world of academic accounting thanks to its diversity of projects, each having its own characteristics. And little by little, he finds his place in the world of research—as strange and new as it is interesting.

His new responsibilities have not entirely taken him from his passion for basketball though. Player, coach, club president—Ed has played all these roles at Martigny Basket, one of the largest youth movements in the canton. Today, he devotes himself to training the youngest age group—a pleasure for him and for them alike.

3 students for the program
Understanding what, and who, is around us

Mathematics and psychology: two scientific fields that one rarely sees sitting side by side on a résumé. But that's exactly how they line up on the CV of Sandrine Tornay, PhD student at Idiap since August 2016.

"Mathematics has always been an obvious choice for me," explains Sandrine Tornay. "And while math is a language via which we can understand the world, psychology is a tool for understanding others."

After obtaining a *Matura* at the College of the Abbey of St-Maurice, the young researcher settles on mathematics—one of the more difficult disciplines to learn unaided—as her chosen field of study and enrolls at EPFL. But the curriculum obliges her to study physics, for which she has little interest. So, her first year over, she moves to the University of Fribourg, adding neurosciences and psychology to her list of studies.

Sandrine is also active in the world of politics. In 2011, she was a candidate in the federal elections under the banner of the PDC's youth branch. "My chances were zero, but I wanted to experience the entire process from the inside," Sandrine explains. And it was during a chance discussion of matters political with Jean-Albert Ferrez that she discovered Idiap. "I asked him if my studies in mathematics could be of interest to the institute. He replied yes, and I did an internship there in the summer of 2012, and then my Master's thesis—both on the theme of speech recognition."

The master key of a university qualification in her pocket, Sandrine opts to teach mathematics at the technical and vocational school in Sion. But, still keen to learn, she eventually returns to Idiap and begins her doctoral studies—on language once more, but this time sign language.

"In collaboration with the Interkantonale Hochschule für Heilpädagogik in Zurich and the University of Surrey in the UK, we are developing an automated assessment system for German sign language—a version of sign language referred to as DSGS," Sandrine explains. "The learner makes a sign in front of a camera, and the machine assesses that sign and helps the signer by commenting on its correctness."

Robots
and mountains

When you have grown up surrounded by mountains and love to climb them, to hike them, or to discover them by mountain bike, Idiap is an ideal place to work. And even if this was not the main reason that prompted Emmanuel Pignat to join the Martigny-based institute, the prospect of being able to work close to those peaks did exert a certain attraction.

Since March 2016, Emmanuel—born in Collombey-Muraz—has worked as a PhD student in Sylvain Calinon's Robot Learning & Interaction group, as part of the European research project I-Dress (see article, page 26). The position is, in fact, the logical continuation of the Master's degree work he carried out at Idiap following the conclusion of his studies in micro-engineering at EPFL.

"I'm passionate about robotics," says Emmanuel. "The fact that a research institute like Idiap is based in a region that brings together my professional interests, hobbies, and family is quite an extraordinary opportunity. So it was with the idea of doing a doctorate here that I decided to join the institute in the context of my Master's studies."

And one year into his thesis work, things are progressing well: "Today, Baxter managed to put a jacket on a patient," Emmanuel explains, "and is already tackling a far more complex task—helping patients put their shoes on."

Between biology
and engineering

Sometimes the aims of a promising candidate turn out not to be an exact match with those of a given research project. And that was the case for Adrian Shajkofci at the end of his Master's studies. But the Valais-Wallis Ambition program solved that problem to everyone's satisfaction.

An amateur pianist in his spare time, Adrian Shajkofci likes to tweak sounds—an artistic pursuit not so far removed from the research that has occupied this Idiap PhD student since September 2016. In Michael Liebling's Computational Bioimaging group, Adrian deals with images; more specifically with microscopic images of fish embryos. "Their hearts beat three to five times per second," the researcher explains. "Visualizing this phenomenon for research purposes—for studying cardiac malformations for example—requires excellent spatial and temporal resolution."

Guaranteeing such high resolution requires mastering both signal acquisition and signal processing. This is why Adrian is working on developing an automated microscopy platform, using artificial intelligence and automatic learning techniques. With a Bachelor's degree in Life Sciences and a Master's degree in Bioengineering—both awarded at EPFL—Adrian is comfortable in his role as a linchpin in an interdisciplinary team of biologists and engineers. "This project is an ideal blend of what I can do and what I like to do," he concludes, "all in the pleasant setting of a close-knit working environment."

43

Awards

Cijo Jose

Internal

Each year Idiap awards prizes to two of its research assistants. The first award is for research, the second for papers published. Candidates for the Research Prize are evaluated, by an internal commission, against five criteria: papers published, teamwork, commitment to a project, communication skills, and autonomy. For the Publication Prize, the institute's leadership draws up an initial list of candidates from papers of which an Idiap research assistant is the primary author. The members of the Advisory Board then evaluate, independently and anonymously, the papers selected.

In 2016, the Research Prize was awarded to Darshan Santani, an independent or exceptionally motivated student whose research focuses on the computational characterization of urban spaces from geo-social network data like hat from Twitter and Foursquare. The Research Publication Prize went to Cijo Jose for his outstanding scientific article entitled "Scalable Metric Learning via Weighted Approximate Rank Component Analysis".

External

This year, Idiap wishes to acknowledge its researchers' brilliant participation at international conferences. The quality of their research was rewarded with several distinctions.

N. LE, D. WU, S. MEIGNIER, AND J.-M.ODOBEZ. MEDIAEVAL 2016, HILVERSUM, OCTOBER 2016, WINNER OF THE PERSON DISCOVERY CHALLENGE
EUMSSI team at the MediaEval Person Discovery Challenge 2016

NIKOLAOS PAPPAS, M. TOPKARA, M. REDI, B. JOU, T. CHEN, H. LIU, AND S.-F. CHANG
ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA RETRIEVAL (ICMR), NEW YORK, JUNE 2016, BEST MULTIMODAL PAPER AWARD
Multilingual Visual Sentiment Concept Matching

Darshan Santani

Theses completed

Six students completed their PhD theses in 2016: Rémi Lebret, Joël Legrand, Dimitri Palaz, Nikolaos Pappas, Darshan Santani, and Raphaël Ullmann.

BUILDING WORD EMBEDDINGS FOR SOLVING NATURAL LANGUAGE PROCESSING

Rémi Philippe LEBRET

September 26, 2016

Thesis directors: Prof. Hervé Bourlard and Dr. Ronan Collobert

Members of the thesis committee: Prof. Jean-Philippe Thiran, Dr. Jean-Cédric Chappelier, Dr. Yves Grandvalet, and Prof. Giuseppe Attardi

WORD SEQUENCE MODELING USING DEEP LEARNING: AN END-TO-END APPROACH AND ITS APPLICATIONS

Joël Yvon Roland LEGRAND

October 28, 2016

Thesis directors: Prof. Hervé Bourlard and Dr. Ronan Collobert

Members of the thesis committee: Dr. Denis Gillet, Dr. Jean-Marc Vesin, Dr. Claire Gardent, and Dr. Cho kyunghyun

TOWARDS END-TO-END SPEECH RECOGNITION

Dimitri Palaz

July 13, 2016

Thesis directors: Prof. Hervé Bourlard and Dr. Ronan Collobert

Members of the thesis committee: Dr. Jean-Marc Vesin, Prof. Jean-Philippe Thiran, Prof. Steve Renals, and Dr. Samy Bengio

LEARNING EXPLAINABLE USER SENTIMENT AND PREFERENCES FOR INFORMATION FILTERING

Nikolaos Pappas

March 4, 2016

Thesis directors: Prof. Hervé Bourlard and Dr. Andrei Popescu-Belis

Members of the thesis committee: Prof. Sabine Süsstrunk, Dr. Jean-Cédric Chappelier, Prof. Thomas Hofmann, and Prof. Shih-Fu Chang

COMPUTATIONAL ANALYSIS OF URBAN PLACES USING MOBILE CROWDSENSING

Darshan Santani

November 17, 2016

Thesis directors: Prof. Daniel Gatica-Perez

Members of the thesis committee: Prof. Pascal Frossard, Prof. Kristof Van Laerhoven, and Dr. Andres Monroy-Hernandez

CAN YOU HEAR ME NOW? AUTOMATIC ASSESSMENT OF BACKGROUND NOISE INTRUSIVE-NESS AND SPEECH INTELLIGIBILITY IN TELECOMMUNICATIONS

Raphaël Marc ULLMANN

June 14, 2016

Thesis director: Prof. Hervé Bourlard

Members of the thesis committee: Prof. Jean-Philippe Thiran, Dr. Hervé Lissek, Prof. Martin Cooke, and Dr. John Beerends
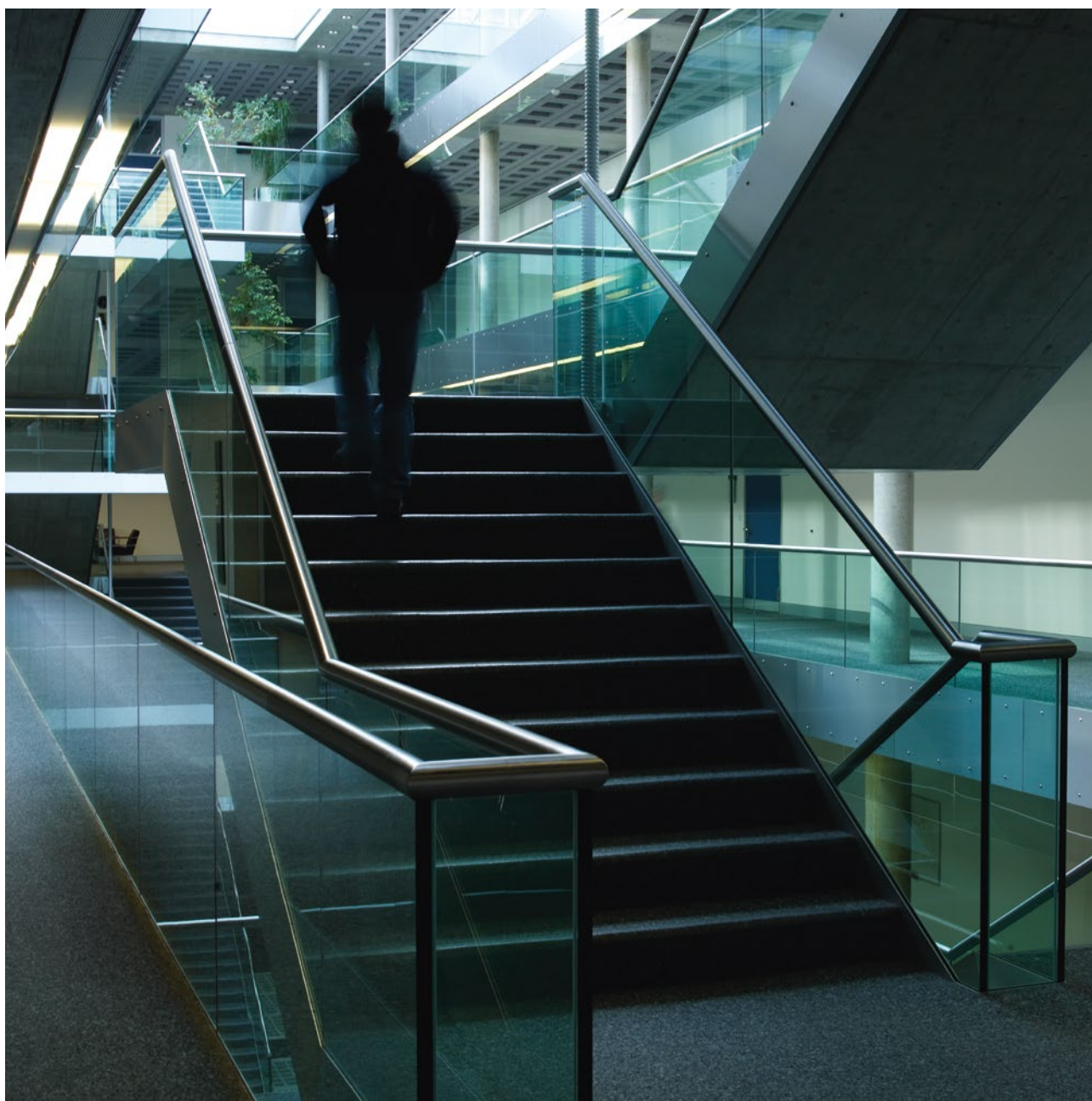
Centre du Parc, Rue Marconi 19,
case postale 592, CH-1920 Martigny

T +41 27 721 77 11
F +41 27 721 77 12
M info@idiap.ch
www.idiap.ch

scientific inserts

I — XLIV

1. Speech and Audio Processing

OVERVIEW

Heads: Prof. Hervé Bourlard (MS and PhD, Polytechnic University, Mons, Belgium, 1982 and 1992), Dr. Philip N. Garner (MEng, University of Southampton, UK, 1991; PhD, University of East Anglia, UK, 2011), Dr. Mathew Magimai-Doss (MS by Research, Indian Institute of Technology Madras, India, 1999; PhD, Ecole Polytechnique Fédérale de Lausanne, Switzerland, 2005), Dr. Petr Motlicek (MS and PhD, Brno University of Technology, Czech Republic, 1999 and 2003).

GROUP OVERVIEW

Speech processing has always been one of the mainstays of Idiap's research portfolio for many years, covering most of the aspects of speech processing such as multilingual automatic speech recognition (ASR), speech synthesis, speech coding, speaker identification, and automatic speech intelligibility evaluation. The expertise and activities of the group encompass statistical automatic speech recognition (based on Hidden Markov Models – HMM, or hybrid systems exploiting connectionist approaches and Deep Neural Networks – DNN), text-to-speech (TTS), speaker recognition (with extensions towards forensics) and generic audio processing (covering sound source localization, microphone arrays, speaker diarization, audio indexing, very low bit-rate speech coding, perceptual background noise analysis for telecommunication systems) and, more recently, Compressive Sensing (CS) and Sparse Recovering theories applied to ASR. The Speech and Audio Processing group in 2016 was composed of 1 head of group, 3 principal investigators, 3 research associates, 6 postdocs, 9 PhD students and 4 interns. In addition to that, the group also worked closely with two Senior Development Engineers from the development team at Idiap on industrial research projects.

KEY SCIENTIFIC OUTPUTS

Idiap is always significantly contributing to both Hidden Markov Model (HMM) and Deep Neural Network (DNN) based approaches applied in acoustic modelling for various speech processing tasks. Use of techniques from HMM and HMM-DNN based speech recognition in HMM and HMM-DNN based speech synthesis resulted in a unified approach to speech recognition and synthesis. The group was well placed to take full advantage of recent advances in "deep" neural networks (DNNs) technology, and the associated Kaldi Automatic Speech Recognition (ASR) toolkit, now used by most of the international speech community. Several key contributions were achieved in speaker recognition tasks, particularly focusing on textdependent speaker verification scenarios and information fusion for large-scale speaker identification. More recently, the group also started exploiting new Compressive Sensing (CS) and Sparse Recovering theories to automatic speech recognition, developing new theoretical links between CS and statistical/HMM-DNN approaches, resulting in improved ASR performance, as well as new spoken term query detection algorithms. Much of the group's recognition has come from contributions to the speech and audio community. These include the AMI corpus[1] used for multiparty conversational speech processing, the Juicer ASR decoder[2], and major contributions to the Kaldi[3] open-source ASR toolkit. More recently, the MediaParl[4] and Walliserdeutsch[5] databases have been released; many other contributions are on the GitHub site[6]. This highlights that, although the group has traditionally worked with English speech, many recent projects are multi-lingual in nature. This has in turn become a focus of the group; one on which we are well placed to capitalise given our geographical location.

---

[1] www.amiproject.org

[2] juicer.amiproject.org/juicer

[3] kaldi-asr.org

[4] www.idiap.ch/dataset/mediaparl

[5] www.idiap.ch/dataset/walliserdeutsch

[6] github.com/idiap

AUTOMATIC SPEECH RECOGNITION

In recent years, our ASR research activities have been expanded from mono-lingual to cross-/multilingual processing. More specifically, in addition to focusing on "majority" languages other than English, such as French and German, Idiap is actively carrying out research in several ASR directions, including:

- **Robust parametrization and acoustic modeling**
  We are still investigating new features (e.g., posterior-based features) and new acoustic models (new forms of Hidden Markov Models, such as KL-HMM, or artificial neural networks) that are more robust to noise and acoustic environments, as well as to speaker variability (e.g., accented speech, or dialect).

- **Cross-lingual and multi-lingual speech recognition**
  Within the scope of the SUMMA project (Figure 4), we focus on investigating and exploiting fast acoustic model adaptation techniques in cross-lingual and multi-lingual scenarios. The resulting speech recogniser relies on a hybrid approach, where an artificial neural network acoustic model is boot-strapped using well-resourced data and adapted to the target language.
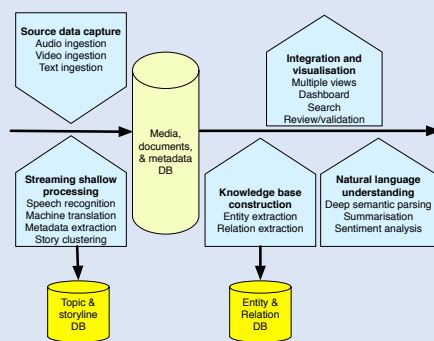
Figure 4

An overview of the SUMMA project, a current focus for multilingual ASR at Idiap.

- **Swiss-languages**
  We continuously improve our speech recognisers for Swiss German and Swiss French and also apply the most recent advances in speech technology employing deep neural networks. The recognisers are also used in commercial products of the Idiap spinoff recapp IT AG. Work on speech recognition for Walliserdeutsch one of the most difficult to understand of the Swiss dialects, was first published in 2014; the associated database is also available for download. Since 2015, we collaborate with recapp IT AG on a wider range of Swiss dialects towards the first commercial product that performs Swiss German (dialect) speech recognition. A new collaboration in 2016 with Swisscom has enabled us to extend our Swiss German work.

- **Template-based ASR**
  Idiap continues to work on template-based or exemplar-based ASR using posterior-based features in the context of projects like SNSF projects A-MUSE, PHASER, and PHASER-QUAD, and CTI project MultiVEO[7]. In this context, there are currently two on-going efforts. The first focuses on development of "sparse" and "low-rank" HMM frameworks by establishing a link between statistical speech recognition formalism and the recent theories of sparse modelling and (probabilistic) principle component analysis (PCA). The framework integrates the advantages of template-based modelling within the HMM sequence matching scheme. The second effort focusses on development of an approach that seamlessly combines both statistical sequence model based ASR system and template-based ASR system into a single ASR system that can yield performance better than the best individual system.

---

[7] www.idiap.ch/en/scientific-research/
projects/MULTIVEO

- **Lexicon development**
  Over the past six years, Idiap has been conducting research on the use of alternative subword units, specifically graphemes, and development of an ASR approach which is capable of handling both acoustic and lexical resource constraints by learning graphemeto- phoneme relationship through the acoustic speech signal (funded through the SNSF-project FlexASR and HASLER Foundation project AddG2SU).

- **Punctuation prediction**
  As our ASR output becomes an input for processing at higher semantic levels, for instance in the SUMMA project, other meta-data such as punctuation becomes necessary. ASR does not normally produce punctuation, but it is possible using similar techniques, notably language modelling and pause detection.

- **Failure analysis of ASR and HSR decoding channels**
  This is a novel strategy developed in 2016 to identify the key sources of information loss in ASR and human speech recognition (HSR). This approach builds on the power of DNN in probabilistic characterization of the subword classes constituting a language. We cast ASR and HSR as separate channels decoding the sequence of sub-word components from their probabilistic representation. Information theoretic measures are developed to assess and quantify the information loss in acoustic modeling for ASR decoding using hidden Markov models.

SPEECH SYNTHESIS AND CODING

- **Text-to-speech synthesis (TTS)**
  Although newer than ASR, TTS is now an established venture for the speech group at Idiap. TTS has been central to several projects, including: SIWIS[8] (Spoken Interaction With Interpretation in Switzerland), D-BOX[9] and SP2[10] (SCOPES project on speech prosody). The TTS work at Idiap was largely aligned with the statistical synthesis trend, which uses the same technology as ASR. However, the group has tracked the recent developments in deep learning which

will dominate future research. SIWIS aimed to do speech to speech translation in Swiss languages, including the concept of accents. A key theme of SIWIS is adaptation of speech prosody, i.e., spoken emotion. This will ensure that, e.g., questions and stressed phrases remain as such after translation. To this end, we have developed novel models of speech prosody. The goal of SIWIS was mirrored in the SP2 project with a focus on eastern-European languages.

- **Speech coding**
  Another research area requiring TTS is speech coding, where very low bit rates can be achieved by using a speech recogniser followed by a synthesiser. Previously, under the RECOD project funded by Armasuisse, the technology lends itself to operate at 200–300 bits per second. The solution relies on deep learning characterization of the phone attribute features dubbed as phonological posteriors. Recent advancement building on the findings of the SNSF project PHASER, led to increased efficiency of the architecture. Unique structures of the phonological posteriors are identified as the sparse pronunciation codes composing natural speech communication; a small size codebook is thus constructed and used for ultra low-bit-rate speech coding. Moreover, work on speech coding continued under the SP2 project, mainly on aspects concerning prosody. This led to the release of the "PhonVoc" toolkit, an end-to-end neural network based phonetic and phonological vocoder.

---

[8] www.idiap.ch/project/siwis

[9] www.idiap.ch/en/scientific-research/projects/DBOX

[10] www.idiap.ch/en/scientific-research/projects/SP2

SPEAKER RECOGNITION AND SPEECH ANALYTICS

In the context of the European SIIP project[11], illustrated by Figure 5, the Speech and Audio Processing group has significantly improved their capabilities in suspect identification applicable to very large scale data. The SIIP technology has successfully passed a proof-of-concept event, demonstrating the performance of the developed technology among the key stakeholders. The developed suspect identification solution can analyse not only lawfully intercepted calls, but also multiple types of social-media channels.

As discussed in Section 2.8, and as part of the SNSF Project UniTS[12], the group is also contributing to the Biometric Person Recognition group, including the development of countermeasures to detect attacks on speaker verification systems through forged speech samples.
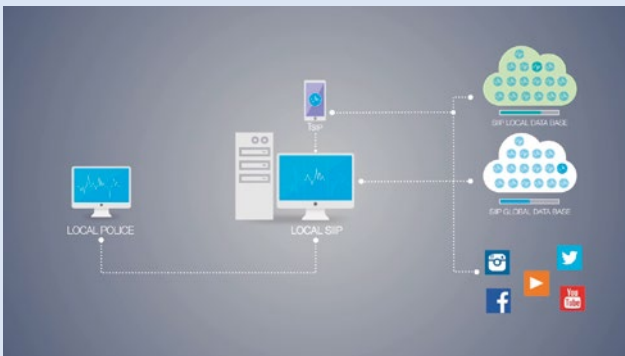
Figure 5

Illustration of SIIP speaker identification system exploiting lawfully intercepted calls,including multiple types of social-media information streams.

FORENSIC SPEAKER VERIFICATION AND AUDIO ANALYSIS

In 2016, Idiap became active in audio forensics which is a field of forensic science for extracting relevant evidences from speech and audio signals that may be ultimately presented as admissible fact in a court of law. Besides a task of enhancement of speech recordings to improve the intelligibility, we have mostly focused on forensic voice comparison to determine the identity of the speaker associated with the strength of evidence. Idiap collaborates with various law enforcement agencies in Switzerland (e.g. Federal and cantonal Police) and abroad to provide direct support in analysing strength of evidences.

LARGE SCALE SPOKEN QUERY RETRIEVAL

In 2016, Idiap became active in audio forensics which is a field of forensic science for extracting relevant evidences from speech and audio signals that may be ultimately presented as admissible fact in a court of law. Besides a task of enhancement of speech recordings to improve the intelligibility, we have mostly focused on forensic voice comparison to determine the identity of the speaker associated with the strength of evidence. Idiap collaborates with various law enforcement agencies in Switzerland (e.g. Federal and cantonal Police) and abroad to provide direct support in analysing strength of evidences.

- **Query-by-example spoken term detection (QbE-STD) based on subspace modelling and detection**
  Idiap continues its research efforts towards keyword spotting and spoken term detection with a focus on searching large audio archives using spoken queries. Over the past 12 months, Idiap has developed pioneering retrieval solutions relying on characterization and detection of the low-dimensional subspace of the DNN phonetic posteriors.

- **Large scale spoken query indexing**
  In 2016, Idiap also developed a powerful methodology for largescale spoken query retrieval relying on hashing.

Contribution of hashing is two-fold: (1) Compressing the information bearing data characteristic and (2) Identifying a data-driven symbolic space for effective search. Idiap hashing solution relies on DNN representation learning at sub-phonetic attribute level to facilitate cross-lingual applications.

OBJECTIVE SPEECH INTELLIGIBILITY ASSESSMENT

Over the past two years, Idiap has been actively developing novel methods to objectively assess intelligibility of speech signal. Specifically, leveraging from our posterior feature based ASR research, we have developed a novel approach where the intelligibility is objectively assessed by estimating a phonetic difference score between reference speech signal and test speech signal. A distinctive advantage of the proposed approach is that the reference speech signal can be replaced by a text transcription of the test speech signal and word level intelligibility assessment can be effectively carried out. In 2016, as part of continuation of armasuisse project OMSI-2015_ARMASUISSE[13], our R&D mainly focused on intelligibility assessment in realistic conditions, such as field operations, narrow band radio transmissions, analog versus digital radio transmissions.

SIGN LANGUAGE RECOGNITION AND ASSESSMENT

In the context of SNSF Sinergia project SMILE[14], Idiap has initiated research on sign language recognition and assessment. The consortium project coordinated by Idiap involves partners from HfH, Zurich and University of Surrey, UK. The end goal of the project is to develop a sign language assessment system that can assist Swiss German sign language learners as well as aid in standardizing a vocabulary production test that can be aligned with levels A1 and A2 of the Common European Framework of Reference for Languages (CEFR).

VI

SOUND LOCALIZATION AND MICROPHONE ARRAY

In 2016, our research activities in the area of microphone array based speech processing and speaker diarization were mainly addressing the problem source localization and reconstruction through binary sparse coding framework.

KEY PUBLICATIONS

[1] A. Asaei, M. Taghizadeh, S. Haghighatshoar, B. Raj, H. Bourlard, V. Cevher, " Binary Sparse Coding of Convolutive Mixtures for Sound Localization and Separation via Spatialization ", *IEEE Transactions on Signal Processing,* Vol. 64, N. 3, pages 567–579, 2016.

[2] M. Cernak, A. Lazaridis, A. Asaei, P. Garner, " Composition of Deep and Spiking Neural Networks for Very Low Bit Rate Speech Coding ", *IEEE/ACM Trans. on Audio, Speech and Language Processing,* pages 2301–2312, 2016.

[3] M. Razavi, R. Rasipuram, M. Magimai.-Doss, " Acoustic data-driven grapheme-to-phoneme conversion in the probabilistic lexical modeling framework ", *Speech Communication,* Vol. 80, pages 1–21, 2016.

[4] I. Himawan, P. Motlicek, D. Imseng, S. Sridharan, " Feature mapping using far-field microphones for distant speech recognition ", *Speech Communication,* Vol. 83, pages 1–9, 2016.

[5] M. Ferràs, S. Madikeri, and H. Bourlard, " Speaker Diarization and Linking of Meeting data ", *IEEE Trans. on Audio, Speech and Language Processing,* 24(11), pages 1935–1945, 2016.

2. Computer Vision and Learning

OVERVIEW

Head: Dr. François Fleuret (MS École Normale Supérieure and University of Paris VI, 1995 ; PhD, University of Paris VI, 2000 ; Habilitation, University of Paris XIII, 2006 ; EPFL MER)

GROUP OVERVIEW

The main objective of the Computer Vision and Learning group is the development of novel machine learning and statistical methods, with an emphasis on their algorithmic efficiency. The application domain we focus on encompasses any processing of images and video streams, such as object detection and scene analysis, tracking of objects and biological structures, and image recognition in general.

The group was composed in 2016 of one head of group, two post-doctoral researchers, four PhD students, two visiting PhD students, one intern, and it had strong interactions with the software development team. Francois Fleuret is also the supervisor of one PhD student from the EPFL Space Engineering Center, and the co-supervisor of two PhD students from the EPFL Computer Vision lab.

KEY SCIENTIFIC OUTPUTS

Our work has resulted in contributions in four different domains: adaptive sampling to deal with very large training sets, large-scale state-of-the-art clustering using geometrical bounds, large-scale metric learning with an orthogonal-inducing regularizer that has a great potential for other learning problems, and mean-field inference combined with deep learning feature extraction for multi-camera object detection.

Additional information and a list of projects are available from www.idiap.ch/cvl.

MACHINE LEARNING

- **Efficient clustering**
  The k-means algorithm is one of the cornerstones of machine learning for real-world applications, in particular for large-scale training. Over the recent years, several approaches have been developed based on the triangular inequality to avoid computing distances when simple bounds insures that it is not necessary.

  In the context of the MASH2 project, we have first improved these exact bounds, and reduced the computational complexity of virtually all state-of-the-art methods, gaining up to a factor two. We also have developed a novel exact algorithm which is the fastest in low dimension. In parallel, we have investigated the "batch" approaches, and proposed a new algorithm combining the triangular inequality and batch of increasing size. It resulted in a very powerful scheme that re-uses computation already done over samples until statistical accuracy requires the use of additional data points.

- **Sub-linear hard sample extraction**
  In the SNSF project DASH, we were interested in speedingup "hard sample" extraction. Most of the state-of-the-art detection algorithms are trained in a sequential manner, improving iteratively the detector by training it with samples miss-classified by its previous version. Since detectors are very accurate, finding a sufficient number of hard samples is computationally intensive, and has traditionally a cost proportional to the number of images to visit.

  Our approach consists in using the prior structure provided for any large collection of images as a recursive decomposition into homogeneous sub-families to concentrate computation in a principled manner. We adapted the Monte-Carlo Tree Search to recursively apply a bandit approach to the exploration-exploitation dilemma.

  In practice, our approach alternates between looking into images it has poor information about, and images from which it has observed a high success rate in finding hard samples.

- **High-dimension similarity measures**
  The WILDTRACK project is a collaboration between Idiap, EPFL and ETHZ around the design of a robust multi-camera pedestrian tracking system.

  One core component of such a system is a re-identification algorithm to automatically associate a person leaving and later re-entering the camera fields of view. This boils down

to learning a metric, so that a small distance between two images is a reliable indicator that the person is the same on both, and large values is a reliable indicator these are different person.

The technique we have developed casts this problem as a rank optimization with a regularizer that enforces the linear projection to remain well-conditioned. Our method allows very large scale learning and beats all existing state-of-the-art methods on standard benchmark data-sets.
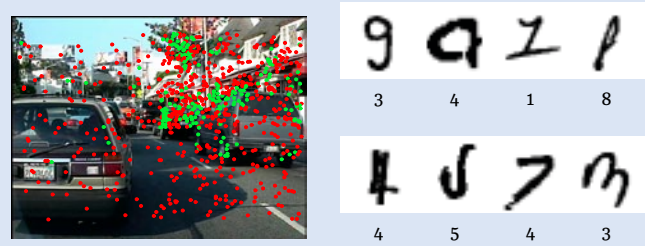
Figure 8
We have developed a novel strategy to find informative training examples in an image that concentrates the computation on difficult images and parts of images to train an object detector (left), and on difficult and unusual examples to train an image classifier (right).
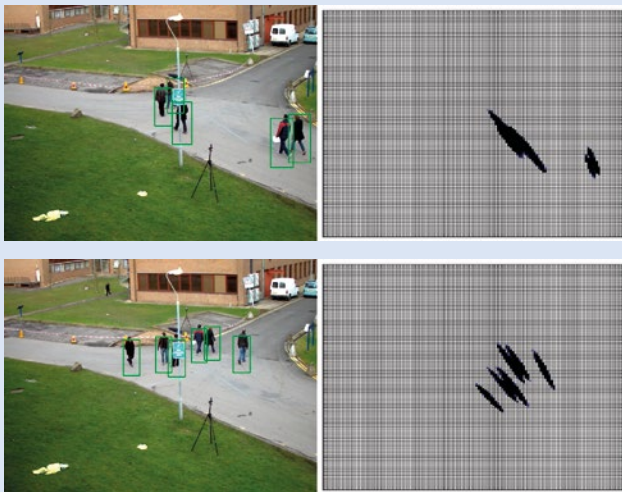
Figure 6
Person detection obtained with a deep neural network, and the corresponding occupancy map estimates.
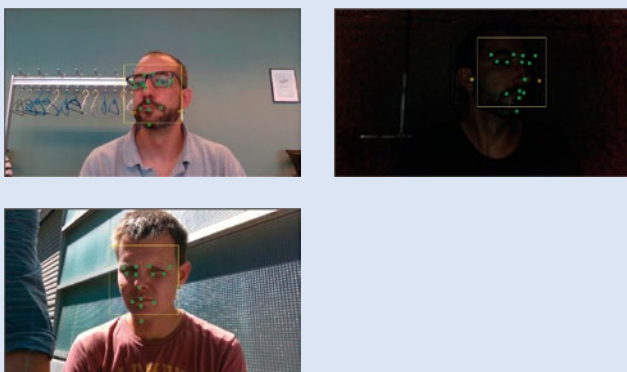
Figure 7
Face alignment examples with qualitative visibility estimation. The color from green to red indicates the visible confidence from high to low.

IMAGE ANALYSIS

- **Depth estimation for planetary surface reconstruction**
  In the framework of the NCCR PlanetS, and in collaboration with the EPFL eSpace center, we are developing a new algorithm to reconstruct the surface of Mars, given images provided by the ExoMars Trace Gas Orbiter.

  State-of-the-art methods use machine learning to automatically match corresponding pairs of images, to estimate the depth of the scene which has been imaged. We improve upon existing technique by relaxing the need for a supervised training set, and using instead crude prior knowledge about the general topography of the planet. This will allow in the future to leverage very large unlabeled training sets.

- **Manufacturing quality control with deep neural networks**
  High-end manufacturing requires accurate quality control, which results in a difficult computer vision problem: while "good" examples are plentiful and quite predictable, the failures are rare and diverse.

  To address the problem, we are investigating generative deep network able to learn from the proper examples alone a sound criterion to detect failures. Our current models are able to decompose an image into two components, one standing for the background image of the materials, and the other for the shape of the manufactured component, which is the only one used to assess the quality of the process.

- **Multi-modal accurate face localization for identification**

  A standard pre-processing used to improve face recognition consists of localizing accurately face landmarks in order to register feature extraction with respect to them. In recent works, we have extended a state-of-art procedure to a multi-modal context where both a rgb image, near-infrared image, and a depth-map are available. Doing so improves drastically the performance, in particular when dealing with poor lighting conditions.

- **Deep learning for multi-camera detection**

  We have a long-lasting collaboration with the CVLab at EPFL around the design of a robust multi-camera tracking system, now in the context of the SNSF WILDTRACK project, in collaboration with ETHZ. We have in particular improved our stateof- the-art mean-field inference scheme Probabilistic Occupancy Map, to leverage deep learning methods for the image feature extraction.

  In parallel, we have developed at Idiap a new approach to adapt a monocular deep learning detector to a multi-camera context. We fine-tune it to the problem of person detection, and then fine-tune a Siamese network–with one such monocular structure per view–on a small multi-view data-set. This staging from a very large generic data-base to a small specific multi-view person data-set allows to beat existing state-of-the art multi-view methods.

- **Advertisement substitution in sport events**

  In a collaboration with ES Concept SA, we have developed a pipeline of algorithms to perform advertisement substitution on-the fly in broadcast video streams. The objective is to deliver a solution able to automatically and accurately localize perimeter advertisement during a sport event, and replace it with another one.

  We have implemented a pipeline which demonstrates such a system, in particular the stability of the estimated board positions in the image plan.

KEY PUBLICATIONS

[1] J. Newling and F. Fleuret. Fast mini-batch k-means by nesting. In Proceedings of the international conference on Neural Information Processing Systems (NIPS), pages 1352–1360, 2016.

[2] C. Jose and F. Fleuret. Scalable Metric Learning via Weighted Approximate Rank Component Analysis. In Proceedings of the European Conference on Computer Vision (ECCV), pages 875–890, 2016.

[3] L. Lefakis and F. Fleuret. Jointly Informative Feature Selection Made Tractable by Gaussian Modeling. Journal of Machine Learning Research (JMLR), 17(182):1–39, 2016.

[4] O. Canévet and F. Fleuret. Large Scale Hard Sample Mining with Monte Carlo Tree Search. In Proceedings of the IEEE international conference on Computer Vision and Pattern Recognition (CVPR), pages 5128–5137, 2016.

[5] X. Wang, E. Turetken, F. Fleuret, and P. Fua. Tracking Interacting Objects Using Intertwined Flows. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 38(11): 2312–2326, 2016.

3. Social Computing

OVERVIEW

Head: Prof. Daniel Gatica-Perez (PhD, University of Washington, USA, 2001; EPFL Adjunct Professor)

GROUP OVERVIEW

Social computing is an interdisciplinary domain that integrates theory and models from ubiquitous computing, social media, machine learning, and social sciences to analyze human and social behavior in everyday life, and to create devices and systems that support social interaction.

The Social Computing group in 2016 was composed of one group head, one research associate, four postdoctoral researchers, four PhD students, and one intern. Research lines investigated in 2016 included: social media analytics and mobile crowdsourcing for cities, ubiquitous conversational interaction analysis, and analysis of Maya hieroglyphic media.

KEY SCIENTIFIC OUTPUTS

Publications on (1) social media analytics and mobile and online crowdsourcing to document youth nightlife patterns and to understand urban perception in cities; (2) multimodal analysis and inference of social constructs in face-to-face employment interviews; and (3) Maya hieroglyphic categorization and retrieval integrating crowdsourcing for data labeling and automatic shape analysis. One EPFL PhD student graduated in 2016 (12 EPFL PhD students graduated since 2002.)

Additional information and a list of projects are available from www.idiap.ch/socialcomputing.

SOCIAL MEDIA ANALYTICS AND MOBILE CROWD-SOURCING IN CITIES

Our work in this domain spans several research lines. First, as part of the SNSF Youth@Night project[15] (A Multidisciplinary Study of Young People's Going Out and Drinking Behaviors, in collaboration with Addiction Switzerland and the University of Zurich), we investigated the use of mobile and online crowdsourcing and social media analytics (Twitter, Instagram, Foursquare) to characterize urban phenomena at scale in cities, both in Switzerland and other world regions[16]. This included the automatic characterization of urban areas according to patterns of youth night activity in a study done for the City of Lausanne[17]; the automatic characterization of physical ambiance from night videos crowdsourced by youth in Zurich and Lausanne (see [1] at the end of this section); and the recognition of social ambiance from social media images of indoor places using deep learning techniques [2] (Figure 9).

In the context of the Civique social innovation platform enabled by the Valais+ project (supported by Loterie Romande)[18], we collaborated with the City of Lausanne to design and implement a mobile survey to collect citizen experiences about street harassment. The results of the study showed the prevalence of this phenomenon, received wide media coverage, and has resulted in concrete city action[19]. This highlights the importance of working jointly with cities to generate insights from data with the participation of city government and citizens. Several other collaborations using the Civique platform are under development.

---

[15] www.youth-night.ch

[16] www.youtube.com/watch?v=71ht15VA0Lw

[17] www.youth-night.ch/static/publications/yatn_rr83.pdf

[18] www.civique.org

[19] www.idiap.ch/en/allnews/results-of-street-harassment-survey-presented-in-lausanne

[20] www.idiap.ch/project/ubimpressed

UBIQUITOUS INTERACTION ANALYTICS

In the context of the SNSF UBImpressed interdisciplinary project[20] (Ubiquitous First Impressions and Ubiquitous Awareness), we have developed computational methods to analyze dyadic interactions in the workplace using sensors (cameras, Kinect devices, microphone arrays, smartphones, and wearables), and automatically infer key social variables (Figure 10). This is joint work with the University of Lausanne and Cornell University. Specifically, we have investigated connections between nonverbal behavior and hirability and other related variables. In all cases, nonverbal cues are automatically measured from audio and video streams, including speaking activity, prosody, and body motion [3].

Figure 10

Sensors available in the social sensing lab. Kinect V2, Microcone, Q-Sensor, Android smartphone, Google Glass.
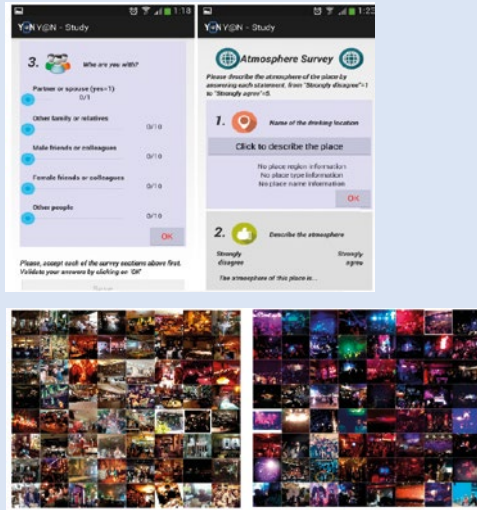

Figure 9

Left: Youth@Night survey logger app. Right: Images classified as Restaurant and Stageby deep network.

This line of work is complemented by the analysis of online video introductions (supported by a project funded by The Ark), where we investigated the issue of interpersonal judgment of dimensions related to impressions of hirability, as recently shown in our work [4].

VISUAL ANALYSIS OF MAYA HIEROGLYPHS

In the context of the SNSF MAAYA project[21] (Multimedia Analysis and Access for Documentation and Decipherment of Maya Epigraphy), we have developed methods for automatic visual analysis of hieroglyphs for the three remaining ancient Maya codices located in museums in Europe (Dresden, Paris, Madrid). Our most recent work has investigated methods for Maya hieroglyphic categorization and retrieval using sparse autoencoders [5] and for glyph stroke segmentation (Figure 11); the use of online crowdsourcing for non-expert glyph labeling, namely localization and segmentation of single signs in glyph-blocks; and graph-based interactive visualization of glyphs according to visual similarity[22]. The visualization work was evaluated with the participation of the general public at two events in 2016: the Idiap Innovation Day (September), and the EPFL Open Days (November).

4. Perception and Activity Understanding

Figure 11
Glyph stroke segmentation, using 500 or 3000 superpixels as spatial support, and two segmentation models (appearance-based method, and when a Conditional Random Field is added).

KEY PUBLICATIONS

[1] J. D. Santani, J.-I. Biel, F. Labhart, J. Truong, S. Landolt, E. Kunstche, and D. Gatica-Perez, The Night is Young: Urban Crowdsourcing of Nightlife Patterns, in Proc. ACM Int. Joint Conf. on Ubiquitous and Pervasive Computing (Ubicomp), Heidelberg, Sep. 2016.

[2] D. Santani, R. Hu, and D. Gatica-Perez, InnerView: Learning Place Ambiance from Social Media Images, in Proc. ACM Int. Conf. on Multimedia (MM), Amsterdam, Oct. 2016.

[3] S. Muralidhar, L. Nguyen, D. Frauendorfer, J.-M. Odobez, M. Schmid Mast, and D. Gatica-Perez, "Training on the Job: Behavioral Analysis of Job Interviews in Hospitality," in Proc. ACM Int. Conf. on Multimodal Interaction (ICMI), Tokyo, Nov. 2016.

[4] L. Nguyen and D. Gatica-Perez, "Hirability in the Wild: Analysis of Online Conversational Video Resumes," IEEE Trans. on Multimedia, Vol. 18, No. 7, pp. 1422–1437, Jul. 2016.

[5] G. Can, J.-M. Odobez, and D. Gatica-Perez, "Evaluating Shape Representations for Maya Glyph Classification," ACM Journal on Computing and Cultural Heritage, Vol. 9, Issue 3, Nov. 2016.

OVERVIEW

Head: Dr. Jean-Marc Odobez (Engineer degree, ENST Bretagne, 1990 ; Ms in Signal Processing, Rennes University, 1990 ; PhD, University of Rennes, France, 1994 ; EPFL MER)

GROUP OVERVIEW

The Perception and Activity Understanding group conducts research in human activities analysis from multi-modal data. This entails the investigation of fundamental tasks like the detection and tracking of people, the estimation of their pose or the detection of non-verbal behaviors, and the temporal interpretation of this information in forms of gestures, activities, behavior or social relationships. These tasks are addressed through the design of principled algorithms extending models from computer vision, multimodal signal processing, and machine learning, in particular probabilistic graphical models and deep learning techniques. Surveillance, traffic and human behavior analysis, human-robot interactions, and multimedia content analysis are the main application domains.

In 2016, the group was composed of one group head, two post-doctoral researchers, three PhD students, one intern, and focused collaborations with the software team.

KEY SCIENTIFIC OUTPUTS

The group is well known for its work on video sequence analysis, probabilistic tracking, non-verbal behavior extraction (in particular attention modeling), and temporal motif discovery. In 2007, the Klewel company was created using its OCR technology. In 2015, the PAU team ranked first at the 2016 MediaEval Person discovery challenge. In 2016, its recent work on 3D face and gaze tracking from cheap RGB-Depth sensors has been patented and has led to the creation of the Eyeware start-up company. Besides the consolidation of his ground

work, the group has investigated deep learning methods for several tasks like gesture recognition, audio-visual speaking activity modeling, gaze, or shape recognition. It has also worked on the integration of its sensing technology into realtime perceptual systems for human-robot interaction (Pepper platform, EU MuMMER project). During the period 2012–2016, the group published 16 journal papers and around 45 conference papers.

Additional information and a list of projects are available from www.idiap.ch/perception.

DEEP LEARNING

In the past years, the team has devoted substantial efforts in investigating the deep learning paradigm to address PAU related tasks. In [2] we studied a multimodal Deep Dynamic Neural Networks (DDNN) framework for the segmentation and recognition of short spontaneous communicative gestures, as illustrated in Fig. 12a. We adopted a semi-supervised hierarchical approach based on Hidden Markov Models learning and fusing emission probabilities from high-level spatio-temporal representations trained using deep neural networks suited to the input modalities: a Gaussian-Bernoulli Deep Belief Network (DBN) to handle the dynamics of skeletal joint information (Fig. 12b), and a 3D Convolutional Neural Network (3DCNN) to manage and fuse batches of depth and RGB images (Fig. 12c). We investigated in [5] deep-temporal representations for the

modeling of audio-visual speaking activities. More precisely, to differentiate dubbing from genuine talking situations, we proposed a method relying on canonical correlation analysis to learn a joint multimodal space, and on long short term memory (LSTM) networks to model cross-modality temporal dependencies linked to anticipation and retention phenomena between the visual and audio streams. The method achieved promising results on challenging data from the EU EUMSSI project.

Other currently addressed tasks include shape representations of Maya Glyph (with Social computing group, see Fig. 11), body-parts localization, gaze estimation, or sound source localization in HRI.

MULTIPLE OBJECT TRACKING

Our previous work resulted in an enhanced Conditional Random Field tracking-by-detection framework with important characteristics making it successful in challenging conditions: long-term pairwise links, pairwise cue measurement costs sensitive to the time interval between two detections, integration of multi-cue association reliability factors, and unsupervised learning avoiding tedious annotations. In 2016, we extended this framework for face track extraction in broadcast data. Thanks to the frame sub-sampling capacity and the learning of program dependent association costs, the model performed better than previous methods with a faster processing, which is useful when processing large video repositories.

GAZE, NON-NERBAL BEHAVIOR EXTRACTION, HUMAN-ROBOT INTERACTIONS

Our long-standing work on NVB perception has been extended in several ways. In the context of the Swiss UBImpressed project, we have worked on robust and accurate head pose tracking from RGB-D data. The approach combines the benefits of the online fitting of a 3D face morphable model with the online 3D reconstruction of the full head. This provides more support when handling extreme head poses, and permits to handle well difficult UBImpressed sequences of role play where a student from the Hospitality industry has to handle a difficult client at a registration desk, as illustrated in Fig. 14. Our long-term research on gaze estimation from cheap RGB-D (depth) cameras like Kinect [1] has been pursued by improving the robustness of our G3E model, which leverages the benefit of the estimation of a geometric model of the eye with the robustness of image driven appearance-based techniques. Figure 13 illustrates the result of applying our tools for gaze coding in multiparty situations.
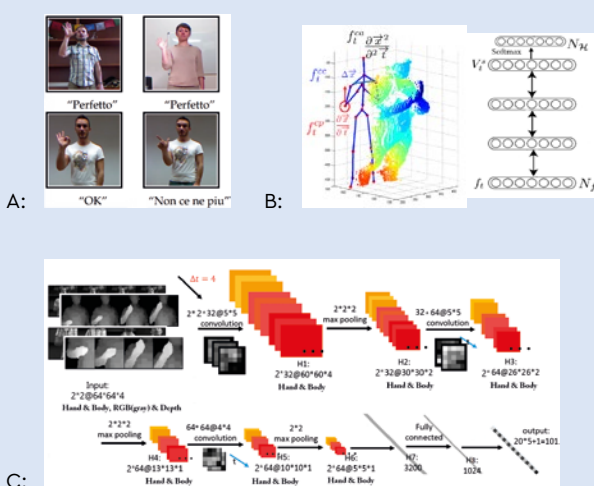
Figure 12

Multimodal gesture recognition from RGB-D data, using deep learning methods. a) Spontaneous communicative gestures to be recognized. b) Encoding of the skeleton information with Restricted Bolzman Machines. c) CNN models of visual and depth hand data.

XIII

In the context of the EU MuMMER project, we have integrated our multi-person tracker on the Pepper robotics platform. This include the extraction of non-verbal cues (attention, head gestures) from both visual only and RGB-D data, and will allow to benchmark our methods in real interactions. In collaboration with KTH, we have worked on the perception of attentiveness from audio-visual (AV) feedbacks [3], investigating which

features of head nods influences attentiveness perception, or whether visual or verbal backchannels or their combination are perceived to be more attentive. Such analysis is useful for endowing robots with the capacity to demonstrate variable level of attentiveness thanks to the synthesis of appropriate subtle AV feedback tokens.

MULTIMODAL FACE AND PERSON DIARIZATION AND NAMING

Identifying people appearing and speaking in multimedia data as we do in the EUMSSI EU project allows to monitor who said what and when and is thus important for the design of search and fast browsing tools of our broadcast programs, as illustrated in Fig. 15. However, person discovery in the absence of prior identity knowledge requires accurate association of visual and auditory cues which in broadcast data faces additional challenges due to narrated voices over muted scenes or dubbing in different languages. In 2016, we have thus investigated dubbing detection (see [5] and deep learning part above), and improved the robustness and efficiency of our video processing pipeline, with an emphasis on the precision of the information delivered to and visualized by the EUMSSI system users.
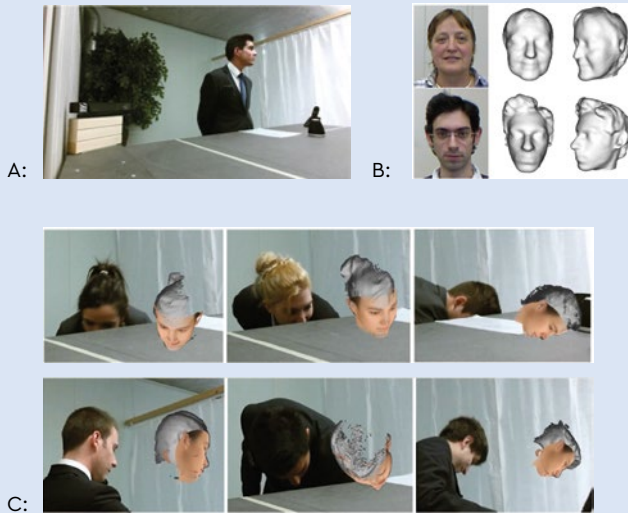

A: B:

C:

XIV

Figure 14

Robust head pose tracking. a) UBImpressed registration desk setup. b) Example of automatically reconstructed heads. c) Example of tracking results in adverse conditions.
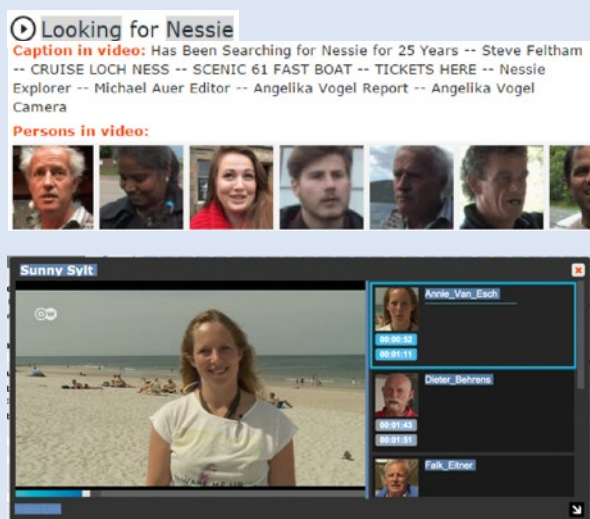

Figure 15

Person identification within the EUMSSI project. Left illustrating the found face clusters of a given Deutsche Welle video document (as well as automatically extracted OCR). Right: fast browsing of documents based on face occurrences.

KEY PUBLICATIONS

[1] K. Funes and J.-M Odobez. Gaze estimation in the 3d space using rgb-d sensors: Towards head-pose and user invariance. Int. Journal of Computer Vision, 118(2):194–216, june 2016.

[2] D. Wu, L. Pigou, P.-J. Kindermans, N. Le, L. Shao, J. Dambre, and J.-M. Odobez. Deep dynamic neural networks for multimodal gesture segmentation and recognition. IEEE Trans. Pattern Anal. Mach. Intell., 38(8):1583–1597, 2016.

[3] C. Oertel, J. Lopes, Y. Yu, K. Funes, J. Gustafson, A. Black, and J.-M. Odobez. Towards building an attentive artificial listener: On the perception of attentiveness in audio-visual feedback tokens. In Int. Conf. on Multimodal Interactions (ICMI), 2016.

[4] Paul Gay, Sylvain Meignier, Paul Deleglise, and Jean-Marc Odobez. CRF-based context modeling for person identification in broadcast videos. Frontiers in ICT: Computer Image Analysis, 3, 2016.

[5] N. Le and J.-M Odobez. Learning multimodal temporal representation for dubbing detection in broadcast media. In ACM Multimedia, Amsterdam, 2016.

5. Robot Learning and Interaction

OVERVIEW

Head: Dr. Sylvain Calinon (MS and PhD, EPFL, 2003 and 2007)

GROUP OVERVIEW

The Robot Learning and Interaction group, created in 2014, focuses on humancentric robotic applications in which the robots can learn new skills by interacting with the end-users. From a machine learning perspective, the challenge is to acquire skills from only few demonstrations and interactions, with strong generalization demands. It requires: 1) the development of intuitive active learning interfaces to acquire meaningful demonstrations; 2) the development of models that can exploit the structure and geometry of the acquired data in an efficient way; 3) the development of adaptive control techniques that can exploit the learned task variations and coordination patterns.

The Robot Learning and Interaction group in 2016 was composed of one postdoctoral fellow, three PhD students, four MSc students and two BSc students.

KEY SCIENTIFIC OUTPUTS

Development of robot learning and control algorithms that can be applied to a wide range of applications, for robots that are either close to us (assistive robots in i-Dress), parts of us (prosthetic hands in TACT-HAND), or far away from us (manipulation skills in deep water with DexROV). Concerned about reproducible research, the group regularly releases open source codes accompanying publications at www.idiap.ch/software/pbdlib/.

Additional information and a list of projects are available from www.idiap.ch/rli.

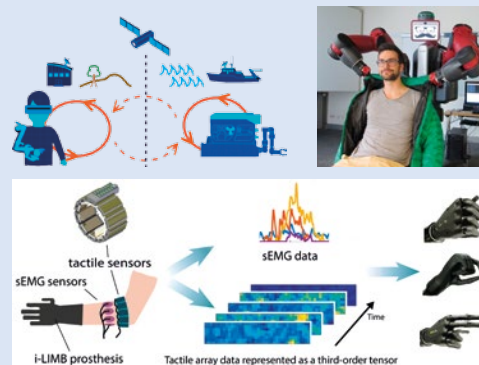## PROBABILISTIC MODELS OF MOVEMENTS AND SKILLS

Figure 16

Human-centric robotic applications. Left: Teleoperation assistance for the control of a bimanual robot (DexROV project), where the robot progressively assists the teleoperator when repetitive or well structured tasks are achieved, as a form of human-robot collaboration. As an example, if the robot has observed that the task of drilling requires the drill to be perpendicular when it approaches a surface, the robot will then automatically orient the drill when it approaches another surface, letting the teleoperator concentrate on the position to drill by delegating the orientation tracking aspect to the robot. Center: Personalized assistance in dressing (i-Dress project), where the robot learns by demonstration the preferences and types of assistance required by each user. This is achieved by extending the movement primitives frequently employed in robotics to a wider repertoire of skills composed of reactive behaviors based on force, impedance, position and orientation. Right: Online learning and adaptive control of a prosthetic hand (TACT-HAND project), where electromyography and tactile sensing data are combined to control a prosthetic hand.

In many robotics applications, demonstrations or experiences are sparse. In such situation, it is important to get as much information as possible from each demonstration. We explore approaches encoding demonstrations from the perspective of multiple coordinate systems. This is achieved by providing a list of observers that could potentially be relevant for the movement or the task to transfer. The collected datapoints are organized in a tensor form (multidimensional arrays). A statistical learning approach is then used to determine the variability and coordination patterns in the movement by considering the different frames of reference simultaneously. This approach allows the orchestration of the different frames to reproduce the movement in a new situation (typically, for adaptation of the movement to new positions of objects).

The proposed task-parameterized model exploits the structure of the task, which can in many robotics problems be expressed in the form of coordinate systems or local projections. It was shown that such approach provides better generalization capability than conventional regression.
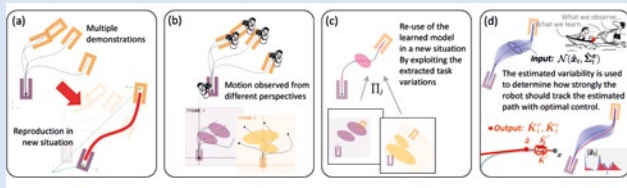
Figure 17

Task-parameterized model to synthesize movements in new situations. (a) Observation of the task in different situations to extract the underlying structure of the task. (b) Probabilistic encoding of movements in multiple coordinate systems. (c) The cross-situational observations of the same task are used to adapt the motion to new situations. (d) Optimal control strategy to reproduce the movement by exploiting the retrieved variability and correlation information.

A DISTINCTIVE VIEW OF OPTIMAL CONTROL

Model predictive control (MPC) and linear quadratic tracking (LQT) are widespread techniques in robot control, but the core formulation of these problems and their associated algorithms can be extended to a wider range of problems, which has often been overlooked in robot learning. In particular, the most simple form of MPC (unconstrained and linear, with a double integrator system) already has advantage for motion synthesis and planning problems, where it can be combined elegantly with probabilistic representations of movements.

This method allows the retrieval of smooth and natural trajectories in a fast and simple manner, by considering variation and coordination constraints. Instead of learning the trajectory directly, the approach allows the learning of the underlying controller to drive the robot. Namely, it learns to reject perturbations only in the directions that would affect task performance (minimal intervention control). This can typically be exploited with torque-controlled robots to regulate the tracking gain and compliance required to reproduce a task in an adaptive manner.

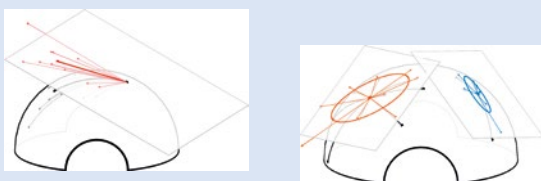## GEOMETRY-AWARE STATISTICAL LEARNING AND

Figure 18

Illustration of iterative gradient descent (left) and Gaussian mixture model (right) on a Riemannian manifold, where the centers of the Gaussians are defined on the manifold, and the covariances are defined in the tangent spaces of the corresponding centers.

REGRESSION

Data in robot learning and control problems have structures and symmetries that are often underexploited. We explore the use of Riemannian manifolds to provide geometry-aware algorithms that can be exploited in several sensing and actuation problems. Examples of Riemannian manifolds in robotics with well known geometries include stiffness and damping gains, inertia, manipulability ellipsoids (symmetric positive definite matrices), orientations (unit quaternions), period movements (phase variable on circle manifold), or rotary joints (e.g., a 2-link planar robot forms a torus manifold).

We explore the use of differential geometry techniques to treat data of various forms in a uniform probabilistic manner (including data in standard Euclidean spaces). In particular, we exploit the potential that several Riemannian manifolds can be combined together to treat problems such as learning/controlling the pose of an end-effector (position and orientation), or learning/controlling both periodic and discrete movements.

It can typically be used to revisit common optimization problems in robotics formulated in standard Euclidean spaces, by removing the constraint of staying on a given manifold, and instead considering an unconstrained problems in spaces that inherently take into account this constraint.

KEY PUBLICATIONS

[1] Tanwani, A.K. and Calinon, S. (2016). Learning Robot Manipulation Tasks with Task-Parameterized Semi-Tied Hidden Semi-Markov Model. IEEE Robotics and Automation Letters (RA-L), 1:1, 235–242.

[2] Calinon, S. (2016). A Tutorial on Task-Parameterized Movement Learning and Retrieval. Intelligent Service Robotics (Springer), 9:1, 1–29.

[3] Rozo, L., Calinon, S., Caldwell, D.G., Jimenez, P. and Torras, C. (2016). Learning Physical Collaborative Robot Behaviors from Human Demonstrations. IEEE Trans. on Robotics, 32:3, 513–527.

[4] Rozo, L., Silvério, J., Calinon, S. and Caldwell, D.G. (2016). Learning Controllers for Reactive and Proactive Behaviors in Human-Robot Collaboration. Frontiers in Robotics and AI, 3:30, 1–11.

[5] Billard, A., Calinon, S. and Dillmann, R. (2016). Learning From Humans. Siciliano, B. and Khatib, O. (eds.). Handbook of Robotics, 2nd Edition, Chapter 74, pp. 1995–2014. Springer.

6. Uncertainty Quantification and Optimal Design

OVERVIEW

Head: PD. Dr. David Ginsbourger (Ph.D. Mines Saint-Etienne 2009, Habilitation Univ. Bern 2014)

GROUP OVERVIEW

The Uncertainty Quantification and Optimal Design group focuses on quantifying and reducing uncertainties in the context of natural and artificial complex systems. Application domains notably include energy and geosciences, with a number of collaborations ranging from safety engineering to hydrology and climate sciences. In all these fields the study of complex systems often relies on expensive data acquisition and model runs, calling for adapted experimental design strategies.

UQOD started at Idiap in September 2015, with members coming from and keeping strong academic ties to the Institute of Mathematical Statistics and Actuarial Science (IMSV) of the University of Bern (UniBE). During the year 2016, the UQOD group has been composed of a permanent senior researcher, a visiting and then Idiap postdoctoral researcher (who now took over a permanent position abroad), a long term visiting IMSV PhD student (now Idiap postdoc), an intern, and three UniBE students occasionally visiting the group (two master students and a PhD student in co-tutelle with Marseille).
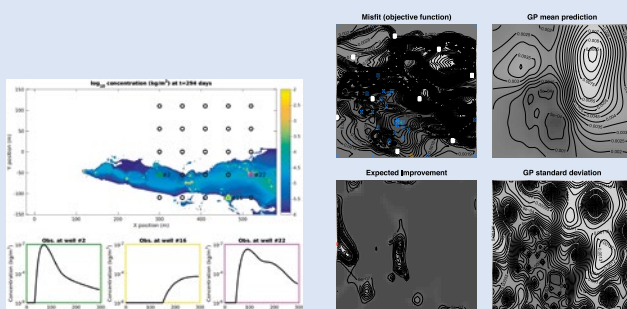
KEY SCIENTIFIC OUTPUTS

Current contributions include efficient algorithms for Bayesian set estimation, notably for estimating and quantifying uncertainties on overcritical parameter regions with Gaussian Process models. Other recent results deal with the interplay between the choice of covariance kernels and properties of Gaussian Process paths, with implications in high-dimensional Gaussian Process modelling and in function prediction under structural constraints. Ongoing work also encompasses novel algorithms for non-stationary modelling of extremes with application in climate sciences, as well as a collaboration with hydrogeologists towards flow simulation-based contaminant source localization.

Additional information and a list of projects are available from www.idiap.ch/uqod.

BAYESIAN OPTIMIZATION AND EMULATION WITH GAUSSIAN PROCESS MODELS

Bayesian global optimization relying on Gaussian Process (GP) emulation has become a standard for optimizing systems which evaluations are prohibitively expensive to evaluate, e.g. with response(s) of interest stemming from heavy numerical simulations. It constitutes one of the core domains of expertise of the UQOD group, with recent methodological contributions ranging from theoretical to methodological questions (such as parallelization, handling large data sets, etc.) and concrete applications. Notably, the group has been involved in a collaboration with researchers in hydrogeology from the universities of Lausanne and Neuchâtel, with the aim to investigate Bayesian optimization for contaminant source localization relying on flow simulations. Figure 19 illustrates some first optimization results based on a fixed geology: Bayesian optimization automatically selects the ground truth contamination source much more efficiently than brute force grid search. This ongoing work has lead to generate objective functions that are to be made publicly available for algorithm benchmarking. The UQOD group also investigates GP emulation per se, with a

Figure 19

Bayesian optimization on a contamination source identification problem.

particular focus on the incorporation of expert knowledge and the identification of structural properties of objective functions through the specification of covariance kernels and the estimation of their parameters. This concerns notably the incorporation of mathematical invariances such as symmetries or harmonicity and also links to global sensitivity analysis, paving the way to novel high-dimensional GP emulation approaches [1, 3].
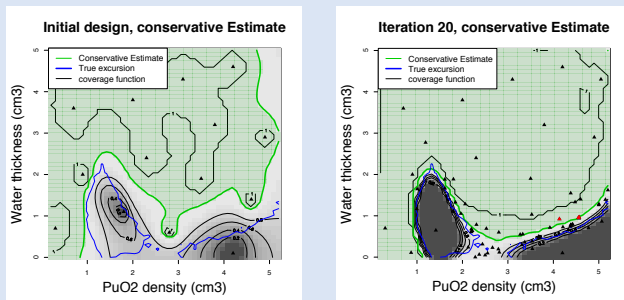

Figure 20

Excursion set estimates obtained on a neutronic criticality safety test case.

COMPUTER EXPERIMENTS FOR THE QUANTIFICATION AND THE REDUCTION OF UNCERTAINTIES

Besides global optimization, the UQOD group has also been focusing on strategies dedicated to other goals such as locating parameter regions leading to a response exceeding a given threshold, that correspond for instance to a dangerous or abnormal behaviour of the considered system. GP-based methods have also been investigated to quantify uncertainties on such sets by using stochastic simulation techniques [2]. Recently, approaches have been proposed for conservative estimation of such sets, i.e. by controlling for instance the probability, under GP assumptions, for set estimates to be completely included in the target set. Figure 20 presents 95% conservative set estimates obtained on a neutronic criticality safety test case (from the French Institut de Radioprotection et de Sûreté Nucléaire) starting from an initial design of experiments (left) and after a novel sequential strategy (right) dedicated to this class of conservative set estimation problem

[5]. This ongoing work, in collaboration with colleagues from Neuchâtel, CentraleSupélec and IRSN, was presented at the SIAM 2016 conference on Uncertainty Quantification. For his PhD thesis entitled " Bayesian Set Estimation Relying on Random Field Priors " defended in November 2016 at the University of Bern, Dario Azzimonti received the C. Moser Award for outstanding Ph.D. or M.Sc. theses of IMSV. In other respects, UQOD has been active in designing computer experiments based on distances between set-valued inputs for output distribution quantization [4].

EXTREME VALUE ANALYSIS FOR NON-STATIONARY TIME SERIES WITH APPLICATION IN CLIMATE SCIENCES

Extreme value analysis is a key approach to a number of phenomena from nature, economic and industry activities. In order to assess risks and make reliable decisions, it is of importance to understand and forecast how extreme some given quantities of interest may be or become, be they for instance climatological or economical variables. The theory of extreme value distributions has been studied for decades, and adequate statistical distributions are known for dealing with maxima and quantify how severe and with what probability extreme events may occur. However, many methods in use require a sample of independent identically distributed maxima in order for the underlying mathematical theory to be valid. This poses serious challenges to climate scientists and beyond, as phenomena exhibit dependencies and also to change over time. In approaches currently being developed within UQOD in collaboration with the Oeschger Center for Climate Change Research we tackle extreme value analysis of non-stationary climatological times series. We investigate algorithms that, in contrast with broadly used block-wise maxima approaches, do not only rely on yearly extremes but also take into account higher frequency information in order to better capture non-stationarities (See Figure 21).
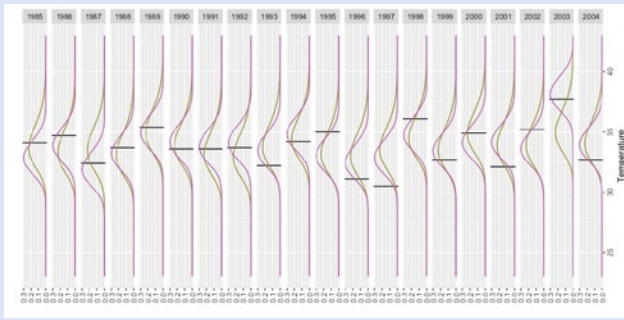
Figure 21

Annual maximum temperatures at Geneva station between 1985 and 2004: observed values (segments) versus retrospectively fitted probability distributions (gray and purple densities).

KEY PUBLICATIONS

[1] On degeneracy and invariances of random fields paths with applications in Gaussian process modelling, David Ginsbourger, Olivier Roustant and Nicolas Durrande, in: Journal of Statistical Planning and Inference, 170:117–128, 2016.

[2] Quantifying uncertainties on excursion sets under a Gaussian random field prior, Dario Azzimonti, Julien Bect, Clément Chevalier and David Ginsbourger, in: SIAM/ASA J. Uncertainty Quantification, 4(1):850- 874, 2016.

[3] On ANOVA Decompositions of Kernels and Gaussian Random Field Paths, David Ginsbourger, Olivier Roustant, Dominic Schuhmacher, Nicolas Durrande and Nicolas Lenz, in: Monte Carlo and Quasi-Monte Carlo Methods, pages 315–330, Springer International Publishing, 2016.

[4] Design of Computer Experiments Using Competing Distances Between Set-Valued Inputs, David Ginsbourger, Jean Baccou, Clément Chevalier and Frédéric Perales, in: mODa 11 – Advances in Model- Oriented Design and Analysis, pages 123–131, Springer International Publishing, 2016.

[5] Contributions to Bayesian set estimation relying on random field priors, Dario Filippo Azzimonti, Inauguraldissertation der Philosophisch-naturwissenschaftlichen Fakultät der Universität Bern, 2016.

7. Computational Bioimaging

OVERVIEW

Head: Dr. Michael Liebling (MS, EPFL, 2000 ; PhD, EPFL 2004 ; postdoc, Caltech, 2004–2007 ; Assistant Prof, University of California Santa Barbara (UCSB), 2007–2013 ; Associate Prof, UCSB, 2013–)

GROUP OVERVIEW

Research in the Computational Bioimaging Group focuses on developing image acquisition, reconstruction and analysis algorithms to study live biological systems. Practical tools aim at (i) extending the physical limits of imaging hardware via techniques including super-resolution and multi-view, space variant deconvolution, and (ii) quantitative analysis of complex biological systems: motion-based image analysis, cell tracking, microscopic fluid flow estimation, and integration of multimodality images.

In 2016, the Computational Bioimaging Group was composed of the head of group, three PhD students and one intern.

KEY SCIENTIFIC OUTPUTS

Recent milestones include the reconstruction of 3D volumes of the beating embryonic heart at frame rates above 1000 volumes per second, temporal super-resolution for sensitive fluorescence cameras, and observation and quantitation of heart development in animal models.

Additional information and a list of projects are available from www.idiap.ch/cbi.

DEVELOPMENT OF OPTICAL PROJECTION TOMOGRAPHY ALGORITHMS TO ENABLE USE OF HIGH-RESOLUTION MICROSCOPY OBJECTIVES

Traditional optical projection tomography requires the use of low numerical aperture objectives in order to achieve a large depth of field and approximate parallel projection geometry. However, low NA objectives suffer from poor resolution, resulting in blur in the reconstructed image. We developed a modified filtered backprojection method suitable for

focal-plane-scanning optical projection tomography (FPS-OPT), where each projection is obtained by scanning through focal planes during collection. We showed that FPS-OPT has an exact inversion formula akin to a filtered backprojection (FBP), but that incorporates the system's point-spread-function to recover a deblurred 3D volume. With simulations, we demonstrated that FPS-OPT permits the use of high numerical aperture objectives that lead to more accurate images. We further illustrated the technique on experimentally acquired data from a fluorescently-labeled zebrafish larva, which shows that our approach reduces out-of-focus blur.
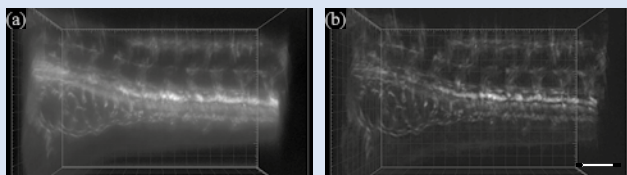
Figure 22

We compare (a) standard FBP with (b) our PSF-aware FBP for 3D fluorescence OPT imaging of a Tg(fli1a:eGFP) zebrafish tail. Scalebar is 100 m. Adapted from [1].

SIMULTANEOUS TEMPORAL SUPERRESOLUTION AND DENOISING FOR CARDIAC FLUORESCENCE MICROSCOPY

Due to low light emission of fluorescent samples, live fluorescence microscopy imposes a tradeoff between spatiotemporal resolution and signal-to-noise ratio. This can result in images and videos containing motion blur or Poisson-type shot noise, depending on the settings used during acquisition. We proposed an algorithm to simultaneously denoise and temporally super-resolve movies of repeating microscopic processes that is compatible with any conventional microscopy setup that can achieve imaging at a rate of at least twice that of the fundamental frequency of the process (above 4 frames per second for a 2 Hz process). Our method combines low temporal resolution frames from multiple cycles of a repeating process to reconstruct a denoised, higher temporal resolution image sequence which is the solution to a linear
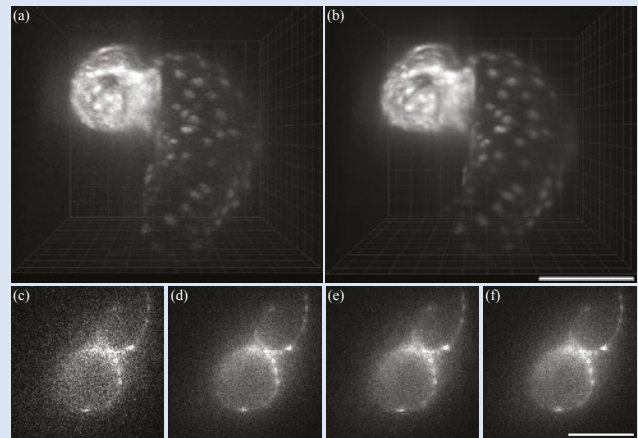

Figure 23

We imaged the heart of a live, 2.5 day old zebrafish embryo imaged in fluorescence at 60 frames per second. (a) low illumination intensity and short integration time used during acquisition result in an image severely corrupted by Poisson-type noise. (b) Our temporal superresolution reconstruction is able to simultaneously temporally superresolve the image sequence and remove much of the noise. Scale bar is 100 m. Adapted from [2].

program that maximizes the consistency of the reconstruction with the measurements, under a regularization constraint. We described, in particular, a parallelizable superresolution reconstruction algorithm and demonstrated its application to live cardiac fluorescence microscopy. Using our method, we experimentally showed temporal resolution improvement by a factor of 1.6, resulting in a visible reduction of motion blur in both on-sample and off-sample frames.

KEY PUBLICATIONS

[1] Kevin G. Chan and Michael Liebling, " A Point-Spread-Function-Aware Filtered Backprojection Algorithm for Focal-Plane-Scanning Optical Projection Tomography," 2016 IEEE International Symposium on Biomedical Imaging, 2016.

[2] Kevin G. Chan, Sebastian J. Streichan, Le A. Trinh and Michael Liebling " Simultaneous temporal superresolution and denoising for cardiac fluorescence microscopy," IEEE Transactions on Computational Imaging, 2016.

8. Biometric Person Recognition

OVERVIEW

Head: Dr. Sébastien Marcel (PhD, University of Rennes, France, 2000 ; Visiting Professor, University of Cagliari, 2010 ; Lecturer, EPFL, 2013-)

GROUP OVERVIEW

The Biometrics group investigates and develops novel image-processing and patternrecognition algorithms for face recognition (2D, 3D, and near-infrared), speaker recognition, vein recognition, anti-spoofing (PAD: presentation attack detection), and emerging biometric modes (EEG and remote photoplethysmography). The group is geared toward reproducible research and technology transfer, using its own signal-processing and machine learning toolbox Bob (www.idiap.ch/software/bob/).

The Biometrics group in 2016 was composed of 1 head of group, 3 research associates, 3 postdocs, 3 PhD students, 1 intern and 1 visitor.

KEY SCIENTIFIC OUTPUTS

The Biometrics group has been pioneering the work on mobile biometrics (face and speaker recognition) and on PAD in face and speaker recognition by sharing the first open databases, organising the first International competitions and producing the first reproducible research studies in the domain. Regarding face PAD in mobile scenario, the group confirmed that the current trend using discriminant classifiers is prone to over-fitting hence resulting in a lack of generalisation on unseen presentation attacks. Regarding voice PAD we demonstrated that the existing methods generalise poorly when different databases or different types of attacks are used for training and testing. These results question the efficiency and practicality of the existing PAD systems, as well as, call for creation of databases with larger variety of realistic speech presentation attacks. The biometrics group also investigated approaches for heterogeneous face recognition and vein recognition.

Additional information and a list of projects are available from www.idiap.ch/biometric.

FACE AND SPEAKER RECOGNITION

We leveraged prior work on distribution modelling for part-based face recognition using session variability modelling techniques. Session variability modelling aims to explicitly model and suppress detrimental within-class (inter-session) variation. We examined two techniques to do this, inter-session variability modelling (ISV) and joint factor analysis (JFA), which were initially developed for speaker recognition. Next, we explored Total Variability modelling (TV), so called i-Vectors originally proposed for speaker recognition, for the task of face recognition. We also developed recently a scalable formulation of Probabilistic Linear Discriminant Analysis (PLDA). PLDA is a probabilistic model that has been shown to provide state-of-the-art performance for both face and speaker recognition.

HETEROGENEOUS FACE RECOGNITION

The task of Heterogeneous Face Recognition (Figure 24) consists in to match face images that were sensed in different modalities, such as sketches to photographs, thermal images to photographs or near infrared to photographs. We proposed a novel and generic approach based on Inter-session Variability Modelling to handle this task.

Figure 24

Illustration of the heterogeneous face recognition problem: matching Visible spectra and Near-Infrared spectra (left), matching Visible spectra and sketch (middle), matching Visible spectra and Thermal spectra (right).

PRESENTATION ATTACK DETECTION

One important aspect of biometric systems is their reliability not only when assaulted by impostors, but also under different types of attacks. One possible security treat is presentation attacks (aka spoofing attacks): an action of outwitting a biometric sensor by presenting a counterfeit biometric evidence of a valid user (Figure 25). It is a direct attack to the sensory input of the biometric system and the attacker does not need previous knowledge about the recognition algorithm. Most of the biometric modalities are not resistant to presentation attacks: a biometric system is usually designed to only recognise identities without concern whether the sample comes from a live person or not. Despite the existence of very sophisticated biometric systems nowadays, the task of implementing presentation attack detection (PAD aka anti-spoofing) schemes for them has attracted much less attention.

Speaker recognition (SR) systems are highly vulnerable to presentation attacks (Figure 26) limiting their wide practical deployment. Therefore, to protect against such attacks, effective PAD techniques, need to be developed. We focused on the integration of PAD and SR systems, demonstrating the effect of score-based integration on recognition and attack detection performance.
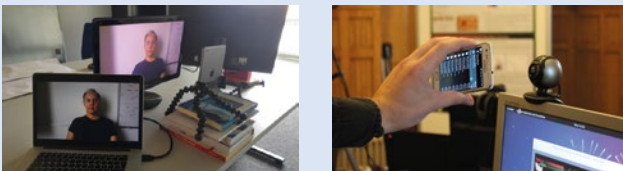
Figure 25

Illustration of video and audio presentation attacks.

REMOTE PHOTOPLETHYSMOGRAPHY

Photoplethysmography (PPG) consists in measuring the variation in volume inside a tissue, using a light source. The aim of remote photoplethysmography (rPPG) is to measure the same variations, but using ambient light instead of structured light and widely available sensors such as a simple webcam (Figure 27).
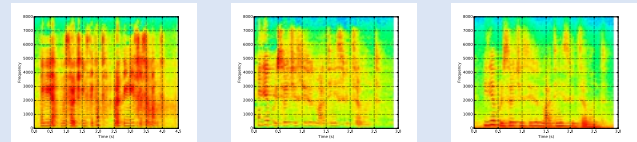
Figure 26

Spectrograms of genuine samples (left) vs. attack (middle and right) samples.

We presented a new, publicly available database containing a relatively large number of subjects recorded under two different lighting conditions. Also, three state-of-the-art rPPG algorithms from the literature were selected, implemented and released as open source free software.

VEIN BIOMETRICS

Vein recognition relies on the complex network of blood vessels located under the human skin. The vascular image of veins, that are located about 3 mm below the surface of the skin, is typically captured under near infra-red (NIR,

Figure 27

Illustration of remote photoplethysmography: colors from the video signal are filtered to produce an estimation of the heart beat signal.

XXII

wavelength between 700 and 1000 nm) illumination. The vein pattern can then be extracted with image pre-process-ing techniques (Figure 28) and used for recognition by any pattern recognition method.

SWISS CENTER FOR BIOMETRICS RESEARCH AND TESTING

In 2014, the Idiap Research Institute launched the " Swiss Center for Biometrics Research and Testing " (www.

Figure 28

Illustration of image processing for vein biometrics (first row for data subject 1 and second row for data subject 2): the original image (left) is geometrically normalised (middle left) and bina-rised with maximum curvature (middle right) or repeated line tracking (right).

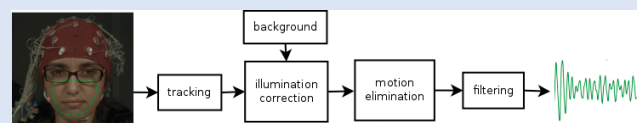biometrics-center.ch), a competence center within the Institute following recent successes in coordinating International research projects in Biometrics (MOBIO, TABULA RASA and BEAT). The aim of this center is to serve as a legacy for these projects and to push for industry-driven research and testing in biometrics. The center attracted the attention of large companies (license, research and testing agreements) and led to many new projects (DARPA, IARPA, CTI).

KEY PUBLICATIONS

[1] P. Korshunov and S. Marcel, " Cross-database evaluation of audio-based spoofing detection systems ", Proceedings of Interspeech, 1705–1709, 2016.

[2] M. Günther and L. El Shafey and S. Marcel, " Face Recognition in Challenging Environments: An Experimental and Reproducible Research Survey ", Face Recognition Across the Imaging Spectrum, T. Bourlai (Eds), pages 247-280, 2016.

[3] I. Chingovska and N. Erdogmus and A. Anjos and S. Marcel, " Face Recognition Systems Under Spoofing Attacks ", Face Recognition Across the Imaging Spectrum, T. Bourlai (Eds), pages 165–194, 2016.

[4] A. Hadid, N. Evans, S. Marcel, and J. Fierrez, " Biometrics systems under spoofing attack: an evaluation methodolo-gy and lessons learned ", IEEE Signal Processing Magazine, 32(5):20–30, 2016.

[5] A. Morales, J. Fierrez, R. Tolosana, J. Ortega-Garcia, J. Galbally, M. Gomez-Barrero,A. Anjos and S. Marcel, " Keystroke Biometrics Ongoing Competition ", IEEE Access, Vol 4, 7736–7746, 2016.

9. Natural Language Processing

OVERVIEW

Head: Dr. Andrei Popescu-Belis (MEng, École Polytechnique, France, 1995 ; MSc, Université Pierre et Marie Curie, France, 1996 ; PhD, Université Paris-Sud, France, 1999)

GROUP OVERVIEW

The Natural Language Processing group studies text analysis at the semantic and pragmatic levels for two main applications: machine translation (MT) and information retrieval (IR). The group aims to improve the state of the art on core semantic and pragmatic analysis problems, such as the disambiguation of nouns and pronouns, topic labeling, keyword extraction, or sentiment analysis – so that they improve in turn MT and IR. Regarding MT, we combine text-level processing techniques with statistical and neural MT systems to improve translation quality. Regarding IR, we design models of document content and sentiment to improve multimedia classification and recommendation.

In 2016, the NLP group had the following members: the head of the group, three postdoctoral students, and two PhD students. A PhD thesis was defended at EPFL in February 2016.

KEY SCIENTIFIC OUTPUTS

In 2016, we have demonstrated that the semantic analysis of noun phrases and pronouns in documents with multiple sentences are beneficial to statistical MT, an innovation that has been met with interest from the community, following our earlier work on discourse connectives and verb tenses. We have shown that a multiple-instance learning framework was appropriate to predict aspect ratings in texts (e.g. sentiments towards aspects of items), and was able to explain the predictions through the weights given to sentences, which

can be used for segmentation and summarization as well. Recently, we designed hierarchical neural networks with attention, enabling cross-lingual transfer for document classification in 8 languages.

Additional information and a list of projects are available from www.idiap.ch/nlp.

DOCUMENT-LEVEL MACHINE TRANSLATION

We focus on a problem that is not often targeted in the current statistical machine translation (SMT) paradigm: the translation of words that depend on long-range relationships between sentences. The main examples are discourse connectives (which have been our main focus in 2010–2014)[23], verb tenses (studied in 2013–2014), and pronouns and noun phrases (our current focus)[24]. The NLP group coordinates a consortium with teams from Geneva, Zürich and Utrecht. In 2016, our main achievements in document-level noun and pronoun MT were the following ones.

We designed a method to enforce the consistency of noun translation, as illustrated in Fig. 29. The method is based on a classifier that decides whether two occurrences of the same noun in a source text should be translated consistently. We combined these classifiers with phrase-based SMT for Chineseto- English and German-to-English, in two ways: automatic post-editing of noun translations vs. reranking of MT hypotheses. Our method closes up to 50% of the gap in BLEU scores between the baseline and the oracle classifier.

In addition to noun consistency, we proposed two methods to improve the translation of pronouns. First, we built a coreference model [1], which captures the probabilistic connection between a pronoun and the features of its antecedent (gender, number and humanness) by learning from the output

---

of an anaphora resolution system on parallel texts. When used in combination with a phrase-based MT decoder, our method improves pronoun translation for English-French and Spanish-English MT. Moreover, in a knowledge-poor approach, we built a pronoun-aware language model that is used to re-rank MT hypotheses [2]. The model estimates the probability of a target pronoun given the gender and number of preceding nouns or pronouns on the target side only. This improved the accuracy of pronoun translation by 5% over the baseline. In the above work, to evaluate the translation of pronouns, we compared them automatically with those in a reference translation by using our APT metric (Accuracy of Pronoun Translation). Using results from the DiscoMT 2015 shared task, we showed that APT correlates well with human judgments.

Figure 29

Inconsistent translations of repeated nouns, in blue, from German (Ex. 1) and Chinese (Ex. 2) into English. Only Ex. 1 is truly mistaken, and is corrected by our MT system.

Finally, we generalized our studies and started to improve the translation of entity mentions (nouns or pronouns) with the help of coreference resolution. We implemented a coreference-aware translation system that helps to disambiguate the translation of mentions by optimizing the similarity of mentiongrouping in source and target documents, either by re-ranking entire sentence, or by post-editing mentions only. Again, this improved Spanish-English translation of pronominal mentions.

TEXT ANALYSIS FOR MULTIMEDIA RECOMMENDATION

We develop methods for text and multimedia information retrieval and recommendation, based on the analysis of content and sentiment, and on the networked structure of (social) media repositories. This work was supported by the SUMMA EU project[25], TheArk Foundation[26], the CTI/KTI[27], and included a collaboration with the EPFL Social Media Lab.

Within a weighted multiple-instance framework, we showed how to jointly learn to focus on relevant parts of documents and to classify them into categories. The model can operate on various sentence and document representations, including intermediate hidden states of a neural network, and has state-of-the-art results for aspect rating prediction (i.e. multi-aspect sentiment analysis). Generalizing our previous work on sentiment analysis [3], we demonstrated the explanatory power of our model by comparing its predicted aspect saliency values with those assigned by humans. To this end, we collected human aspect values over a set of audiobook reviews, and made them available as the HATDOC dataset [4]. Moreover, we showed that the aspect saliencies benefit text segmentation and summarization. We also showed that this model is mathematically equivalent to attention mechanisms recently proposed for neural networks.

We developed a state-of-the-art architecture for multilingual document modeling, based on hierarchical neural networks. We explored 3 options for multilingual training over disjoint label spaces: sharing the parameters at each layer of the network, sharing the attention mechanisms at each layer, and sharing both. We demonstrated that these architectures are useful for transferring knowledge to low-resource languages, and that they also improve over monolingual models on the full-scale data. The models were applied to topic labeling in the SUMMA project, using 600k articles from Deutsche Welle,

XXV

in 8 languages, with multiple topic labels. We thus delivered a deep multilingual tagging library for the SUMMA demonstrator, illustrated in Fig. 30 for one test article.
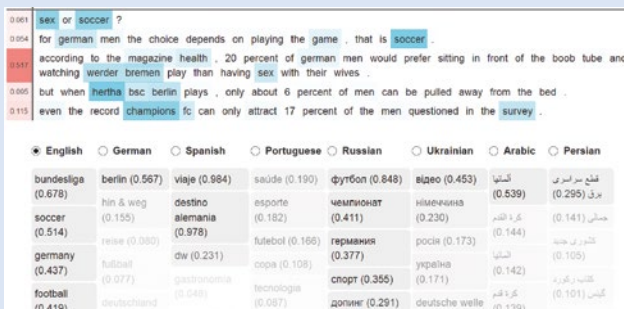
Figure 30

Multilingual topic labeling in 8 languages of a news article from Deutsche Welle.

Working on personalized search for information watch with Faveeo SA, we designed solutions for two recommendation tasks. First, we designed a method to suggest additional search terms to users, related to their initial query terms but diversifying them too. Our comparative study of two word representations, probabilistic topic models vs. word embeddings, showed the superiority of the latter [5]. Moreover, we designed a document recommender, which extends the initial set of results from Faveeo based on what users 'liked', using low-dimensional vector representations of documents.

Within two focused collaborations, we applied our know-how of text similarity methods to two problems. We designed for Unono.net modules for recommending news items to users of their portal, and for recommending profiles (i.e. CVs of applicants) to employers based on their job descriptions, and viceversa. In collaboration with the EPFL Social Media Lab, we designed a module for extracting and clustering n-grams from social media, and visualizing the results. Exploring results under a range of parameters, our tools helped to observe the propagation of opinions about 6 French TV series in social media.

KEY PUBLICATIONS

[1] Luong N.Q. & Popescu-Belis A. (2016) – Improving Pronoun Translation by Modeling Coreference Uncertainty. Proceedings of WMT 2016 (First Conference on Machine Translation), Research Papers, Berlin, Germany, p. 12–20.

[2] Luong N.Q. & Popescu-Belis A. (2016) – A Contextual Language Model to Improve Machine Translation of Pronouns by Re-ranking Translation Hypotheses. Proceedings of EAMT 2016 (19th Annual Conference of the European Association for Machine Translation), Riga, Latvia, p. 292–304.

[3] Pappas N. & Popescu-Belis A. (2016) – Human versus Machine Attention in Document Classification: A Dataset with Crowdsourced Annotations. Proceedings of the EMNLP SocialNLP workshop (4th Int. Workshop on Natural Language Processing for Social Media), Austin, TX, p. 94–100.

[4] Pappas N. & Popescu-Belis A. (2016) – Adaptive Sentiment-Aware One-Class Collaborative Filtering. Expert Systems With Applications, vol. 43, p. 23–41.

[5] Habibi M., Mahdabi P. & Popescu-Belis A. (2016) – Question Answering in Conversations: Query Refinement Using Contextual and Semantic Information. Data & Knowledge Engineering, vol. 106, p. 38–51.

Projects in progress during 2016

PROJECTS FUNDED BY EUROPEAN AND INTERNATIONAL AGENCIES

Name	**MACADAMS** (Modifying Adhoc Centralised Advertisement with Digit Arena Multicast over Satellite)
Funding	Eurostars Programme
Coordinator	Digit Arena
Duration	2016.09.01 – 2018.08.30
Partner(s)	Idiap Research Institute, Ecole Centrale de Lyon, European Broadcast Union

Name	**BEAT** (Biometrics Evaluation and Testing)
Funding	FP7 – STREP – SEC
Coordinator	Idiap Research Institute
Duration	2012.03.01 – 2016.02.28
Partner(s)	École Polytechnique Fédérale de Lausanne, University of Surrey, Commissariat à l'Énergie Atomique, Universidad Autónoma de Madrid, Katholieke Universiteit Leuven, Sagem Securité SA, TUBITAK Informatics and Information Security, Advanced Technologies Research Center, TÜViT

Name	**EUMSSI** (Event Understanding through Multimodal Social Stream Interpretation)
Funding	FP7 – IP – ICT
Coordinator	Fundacio Barcelona Media
Duration	2013.12.01 – 2016.11.30
Partner(s)	Idiap Research Institute, Université du Maine, Gottfried Wilhelm Leibniz Universität Hannover, Gesellschaft zur Förderung der angewandten Informationsforschung EV, Deutsche Welle, Video Stream Networks SL

Name	**SIIP** (Speaker Identification Integrated Project)
Funding	FP7 – SEC – 2013 – 1
Coordinator	Verint System Ltd
Duration	2014.05.01 – 2018.04.30
Partner(s)	Idiap Research Institute, Sail Labs Technology Ag, Singularlogic Anonymos Etairia Pliroforiakon Systimaton & Efarmogon Pliroforikis, Green Fusion Limited (Data Fusion International), Synthema S.R.L., Ok2Go, Loquendo Spa, International Biometric Group (Uk) Limited, Cassidian Sas, Rijksuniversiteit Groningen, Inov Inesc Inovacao – Instituto De Novas Tecnologias, University OfWarwick, Laboratorio Di Scienze Della Cittadinanza, The International Criminal Police Organization, Police Service Of Northern Ireland, Ministério Da Justiça, Lisboa

Name	**TESLA** (An Adaptive Trust-based e-assesment System for Learning)
Funding	H2020 – IA – ICT
Coordinator	Fundacio per a La Universitat Oberta de Catalunya
Duration	2016.01.01 – 2018.12.31
Partner(s)	Idiap Research Institute, Imperial College London, The Open University, Télécom Sudparis, Open Universiteit Nederland, European Association For Quality Assurance In Higher Education Aisbl, Universite De Namur Asbl, Agència Per A La Qualitat Del Sistema Universitari De Catalunya, Lplus Gmbh, Sofiiski Universitet Sveti Kliment Ohridski, Protos Sistemas De Información, S.L., Technical University Of Sofia, Anadolu University, Jyvaskylan Yliopisto, European Quality Assurance Network For Informatics Education E.V., Instituto Nacional De Astrofisica, Optica Y Electronica, Wfsw, Sa, Institut Mines-Telecom

ACTIVE AND GRANTED PROJECTS IN 2016

Name **MUMMER** (MultiModal Mall Entertainment Robot)
Funding H2020 – RIA – ICT
Coordinator University of Glasgow
Duration 2016.03.01 – 2020.02.28
Partner(s) Idiap Research Institute, Centre national de la recherche scientifique, Aldebarab Robotics, Teknologian Tutkimuskekus VTT, Kiinteistö Oy Ideapark AB

Name **SUMMA** (Scalable Understanding of Multilingual Media)
Funding H2020 – RIA – ICT
Coordinator University of Edinburgh
Duration 2016.02.01 – 2019.01.31
Partner(s) Idiap Research Institute, University College London, British Broadcasting Corporation, Deutsche Welle, Priberam Informatica S.A., Leta, Qatar Computing Research Institute

Name **4DHEART** (4D analysis of heart development and regeneration using advanced light microscopy)
Funding H2020 – MSCA – ITN
Coordinator Fundacion Centro Nacional de Investigaciones Cardiovasculares Carlos III
Duration 2016.10.01 – 2020.09.30
Partner(s) Idiap Research Institute, Centre national de la recherche scientifique, Universität Bern, Acquifer AG, Bitplane AG, Leica Microsystems CMS GmbH, 4D-Nature Imaging Consulting, S. L., Centre Européen de Recherche en Biologie et Médecine

Name **DEXROV** (Effective Dexterous ROV Operations in Presence of Communications Latencies)
Funding H2020 – RIA – BG
Coordinator Space Applications Services
Duration 2015.03.01 – 2018.08.31
Partner(s) Idiap Research Institute, Jacobs University Bremen, COMEX SA., Interuniversity Center of Integrated Systems for the Marine Environment, Graal Tech s.r.l.

Name **MALORCA** (Machine Learning of Speech Recognition Models for Controller Assistance)
Funding H2020 – SESAR – RIA
Coordinator Deutsches Zentrum für Luft- und Raumfahrt EV
Duration 2016.04.01 – 2018.03.31
Partner(s) Idiap Research Institute, Universität des Saarlandes, Rízení letového provozu Ceské republiky, Austro Control Österreichische Gesellschaft für Zivilluftfahrt MBH

Name **SWAN** (Secure Access Control over Wide Area Network)
Funding Research Council of Norway ICT initiative
Coordinator Hogskolen I Gjovik
Duration 2016.01.01 – 2018.12.31
Partner(s) Idiap Research Institute, MORPHO, Bankenverband, Universitetet Oslo, Zwipe AS

Name **SAVI** (Media Forensics)
Funding DARPA
Coordinator SRI International
Duration 2016.05.19 – 2020.05.18
Partner(s) Idiap Research Institute

PROJECTS FUNDED BY SWISS AGENCIES

Name	**25ᴱ** (Célébrations du 25ᵉ anniversaire de l'Idiap)
Funding	Loterie Romande
Coordinator	Idiap Research Institute
Duration	2016.09.01 – 2016.09.30

Name	**VALAIS+** (Une plateforme pour mieux connaî-tre l'espace de vie du canton)
Funding	Loterie Romande
Coordinator	Idiap Research Institute
Duration	2015.09.01 – 2017.03.31

Name	**LIVEHEART** (The Cellular Basis of Cardiac Development Revealead by Live Imaging)
Funding	SNSF – ANR
Coordinator	Institut de génétique et de biologie molécu-laire et cellulaire
Duration	2016.03.01 – 2019.02.28
Partner(s)	Idiap Research Institute, Universität Bern, École Polytechnique Paris

Name	**ODESSA** (Online Diarization Enhanced by recent Speaker identification and Sequential learning Approaches)
Funding	SNSF – ANR
Coordinator	Centre national de la recherche scientifique
Duration	2016.03.01 – 2019.02.28
Partner(s)	Idiap Research Institute, EURECOM

Name	**TACT-HAND** (Improving control of prosthetic hands using tactile sensors and realistic ma-chine learning)
Funding	SNSF – DACH
Coordinator	Deutsches Zentrum für Luft- und Raumfahrt EV
Duration	2016.04.01 – 2019.03.31
Partner(s)	Idiap Research Institute, Universität Bielfeld

Name	**MAAYA** (Multimedia Analysis and Access for Documentation and Decipherment of Maya Epigraphy)
Funding	SNSF – DACH
Coordinator	Idiap Research Institute
Duration	2013.09.01 – 2017.04.30
Partner(s)	University of Geneva, University of Bonn

Name	**PHASER-QUAD** (Parsimonious Hierarchical Automatic Speech Recognition and Query Detection)
Funding	SNSF – Division II
Coordinator	Idiap Research Institute
Duration	2016.10.01 – 2018.09.30

Name	**COMETS-M** (Computational Methods for Temporal Super-resolution Microscopy)
Funding	SNSF – Division II
Coordinator	Idiap Research Institute
Duration	2016.04.01 – 2019.03.30

Name	**HFACE** (Heterogeneous Face Recognition)
Funding	SNSF – Division II
Coordinator	Idiap Research Institute
Duration	2014.07.01 – 2018.06.30

Name	**UNITS** (Unified Speech Processing Framework for Trustworthy Speaker Recognition)
Funding	SNSF – Division II
Coordinator	Idiap Research Institute
Duration	2015.07.01 – 2018.06.30

Name	**A-MUSE** (Adaptive Multilingual Speech Processing)
Funding	SNSF – Division II
Coordinator	Idiap Research Institute
Duration	2012.10.01 – 2016.09.30

ACTIVE AND GRANTED PROJECTS IN 2016

Name	**DASH** (Object Detection with Active Sample Harvesting)
Funding	SNSF – Division II
Coordinator	Idiap Research Institute
Duration	2012.10.01 – 2016.09.30

Name	**PHASER** (Parsimonious Hierarchical Automatic Speech Recognition)
Funding	SNSF – Division II
Coordinator	Idiap Research Institute
Duration	2014.06.01 – 2016.05.31

Name	**I-DRESS** (Assistive Interactive robotic system for support in DRESSing)
Funding	SNSF – ERA-NET Chist Era
Coordinator	Institut de Robotica i Informatica Industrial, CSIC-UPC
Duration	2015.12.01 – 2018.11.30
Partner(s)	Idiap Research Institute, University of the West of England

Name	**YOUTH@NIGHT** (A multi-disciplinary multi-method study of young people's outgoing and drinking behaviors)
Funding	SNSF – Interdisciplinary project
Coordinator	Sucht Schweiz – Research Institute
Duration	2014.01.01 – 2017.12.31
Partner(s)	Idiap Research Institute, Universität Zürich

Name	**PLATFORM-MMD** (Platform for Reproducible Acquisition, Processing, and Sharing of Dynamic, Multi-Modal Data)
Funding	SNSF – R'equip
Coordinator	Idiap Research Institute
Duration	2016.07.01 – 2017.06.30

Name	**SP2** (Speech Prosody)
Funding	SNSF – Scopes
Coordinator	Idiap Research Institute
Duration	2014.04.01 – 2016.05.31
Partner(s)	Budapest University of Technology and Econom, Ss. Cyril and Methodius University, Universizy of Novi Sad

Name	**SMILE** (Scalable Multimodal sign language Technology for sIgn language Learning and assessmEnt)
Funding	SNSF – Sinergia
Coordinator	Idiap Research Institute
Duration	2016.03.01 – 2019.02.28
Partner(s)	University of Surrey, University of Applied Sciences of Special Needs Education

Name	**SIWIS** (Spoken Interaction with Interpretation in Switzerland)
Funding	SNSF – Sinergia
Coordinator	Idiap Research Institute
Duration	2012.12.01 – 2016.11.30
Partner(s)	University of Edinburgh, University of Geneva, Eidgenössische Technische Hochschule Zürich

Name	**MODERN** (Modeling discourse entities and relations for coherent machine translation)
Funding	SNSF – Sinergia
Coordinator	Idiap Research Institute
Duration	2013.08.01 – 2017.07.31
Partner(s)	University of Geneva, Universiteit Utrecht, Universität Zürich

xxx

Name UBIMPRESSED (Ubiquitous First Impressions and Ubiquitous Awareness)
Funding SNSF – Sinergia
Coordinator Idiap Research Institute
Duration 2014.01.01 – 2017.06.30
Partner(s) Université de Neuchâtel, Cornell University

Name WILDTRACK (Tracking in the Wild)
Funding SNSF – Sinergia
Coordinator Idiap Research Institute
Duration 2014.01.01 – 2017.12.31
Partner(s) École Polytechnique Fédérale de Lausanne, Eidgenössische Technische Hochschule Zürich

Name OMSI-EXT (Objective Measurement of Speech Intelligibility)
Funding ArmaSuisse
Coordinator Idiap Research Institute
Duration 2016.04.01 – 2016.12.31

Name MIRROR (Mobile Data to Improve Decision Making in African Cities)
Funding EPFL collaboration
Coordinator Idiap Research Institute
Duration 2014.01.01 – 2016.12.31
Partner(s) École Polytechnique Fédérale de Lausanne

Name CREM-IDIAP (Pour une recherche fondamentale et appliquée au service des systèmes énergétiques territoriaux en Valais)
Funding Etat du Valais
Coordinator Centre de recherches énergétiques et municipales
Duration 2016.11.01 – 2017.03.31
Partner(s) Idiap Research Institute

Name SETVISU (Learning and visualizing dangerous regions in multivariate parameter spaces)
Funding Hasler Foundation
Coordinator Idiap Research Institute
Duration 2016.12.01 – 2017.08.31

Name L-PASS (Linguistic-Paralinguistic Speech Synthesis)
Funding Hasler Foundation
Coordinator Idiap Research Institute
Duration 2015.12.01 – 2016.02.28

Name RECAPP (Making speech technology accessible to Swiss people)
Funding Hasler Foundation
Coordinator Idiap Research Institute
Duration 2015.01.01 – 2016.12.31

Name MASH-2 (Massive Sets of Heuristics for Machine Learning II)
Funding Hasler Foundation
Coordinator Idiap Research Institute
Duration 2013.07.01 – 2017.06.30

Name ADDG2SU (Flexible Acoustic Data-Driven Grapheme to Subword Unit Conversion)
Funding Hasler Foundation
Coordinator Idiap Research Institute
Duration 2013.03.01 – 2017.02.28

Name SODS (Semantically Self-Organized Distributed Web Search)
Funding Hasler Foundation
Coordinator Idiap Research Institute
Duration 2012.01.01 – 2016.04.30

PROJECTS FUNDED BY INDUSTRIAL PARTNERS

Name **3DFINGERVEIN** (3D FingerVein Biometrics)
Funding CTI
Coordinator HES-SO Valais
Duration 2016.06.01 – 2018.03.31
Partner(s) Idiap Research Institute, ITServices Sonna Sàrl

Name **BIOWAVE** (A BIOmetric Watch Activated by VEins)
Funding CTI
Coordinator Centre Suisse d'Electronique et de Microtechnique
Duration 2016.02.01 – 2017.07.31
Partner(s) Idiap Research Institute, Biowatch SA

Name **ESGEM** (Enhanced Swiss German mEdia Monitoring)
Funding CTI
Coordinator Idiap Research Institute
Duration 2016.06.01 – 2018.01.31
Partner(s) recapp, ARGUS der Presse AG

Name **FARGO** (Convenient and Secure 3D Face Recognition based on RGB-D Cameras)
Funding CTI
Coordinator Idiap Research Institute
Duration 2016.05.01 – 2017.10.31
Partner(s) KeyLemon SA

Name **IMIM** (Intelligent Monitoring for In-line Manufacturing)
Funding CTI
Coordinator AISA Automation Industrielle SA
Duration 2016.07.01 – 2018.02.28
Partner(s) Idiap Research Institute

Name **VIEW-2** (Visibility Improvement for Events Webcasting)
Funding CTI
Coordinator HES-SO Valais
Duration 2016.08.01 – 2018.03.31
Partner(s) Idiap Research Institute, Klewel

Name **MOGS** (2e génération de l'app MOGS)
Funding CTI
Coordinator Mogs SA
Duration 2016.03.01 – 2016.05.31
Partner(s) Idiap Research Institute

Name **SWISKO** (Swiss-Korean project to develop and integrate new wearable sensors into the existing DomoSafety ambient sensor system.)
Funding CTI
Coordinator Idiap Research Institute
Duration 2015.12.01 – 2017.07.31
Partner(s) DomoSafety S.A., Universität Bern, HES-SO Vaud

Name **SCOREL2** (Automatic scoring and adaptive pedagogy for oral language learning)
Funding CTI
Coordinator Idiap Research Institute
Duration 2014.06.01 – 2016.01.31
Partner(s) SpeedLingua

Name **FAVEO** (Accelerating online information discovery through context-driven and behaviour-based personalization of search)
Funding CTI
Coordinator Idiap Research Institute
Duration 2014.09.01 – 2016.03.31
Partner(s) École Polytechnique Fédérale de Lausanne, Faveeo

ACTIVE AND GRANTED PROJECTS IN 2016

Name	**VIDEOPROTECTOR** (Morphean VideoProtector)
Funding	CTI
Coordinator	Idiap Research Institute
Duration	2014.06.01 – 2016.01.31
Partner(s)	HES-SO Fribourg

Name	**DIGIT-ARENA** (Real-Time Perimeter Board Content Digital Replacement)
Funding	CTI
Coordinator	Idiap Research Institute
Duration	2014.12.01 – 2016.07.30
Partner(s)	Digit Arena

Name	**ELEARNING-VALAIS 3.0** (eLearning-Valais 3.0)
Funding	Fondation The Ark
Coordinator	Formation universitaire à Distance
Duration	2016.03.01 – 2018.01.31
Partner(s)	Idiap Research Institute, Klewel

Name	**LIFE** (Patrouille des Glaciers – fatigue)
Funding	Fondation The Ark
Coordinator	Clinique romande de réadaptation
Duration	2016.08.01 – 2017.03.31
Partner(s)	Idiap Research Institute, Life

Name	**AUDIOSEARCHDEMO** (Demonstration for AudioSearch)
Funding	Fondation The Ark
Coordinator	Idiap Research Institute
Duration	2016.05.01 – 2016.07.31

Name	**EYEWARE** (Eyeware)
Funding	Fondation The Ark
Coordinator	Idiap Research Institute
Duration	2015.09.01 – 2016.04.30

Name	**HYBRID** (Hybrid Recommender System)
Funding	Fondation The Ark
Coordinator	Unono
Duration	2015.08.01 – 2016.03.31
Partner(s)	Idiap Research Institute

Name	**PUNK** (Punktuation)
Funding	Fondation The Ark
Coordinator	recapp
Duration	2015.05.01 – 2016.04.30
Partner(s)	Idiap Research Institute

Name	**SUVA** (Recomed, Intégration de la transcription vocale dans le dossier patient informatisé CRR)
Funding	Fondation The Ark
Coordinator	Idiap Research Institute
Duration	2014.03.01 – 2016.01.31
Partner(s)	Cimark, Clinique romande de réadaptation

Name	**NMTBENCHMARK** (Training and Benchmarking Neural MT and ASR Systems for Swiss Languages)
Funding	Industrial Project
Coordinator	Idiap Research Institute
Duration	2016.12.01 – 2017.11.30
Partner(s)	École Polytechnique Fédérale de Lausanne

Name	**SNACK** (Understanding how to balance the enjoyment of food with the effects of calories)
Funding	Industrial Project
Coordinator	École Polytechnique Fédérale de Lausanne
Duration	2016.03.01 – 2018.02.28
Partner(s)	Idiap Research Institute

Name	**DRACULA** (Detect and track people/object in order to deliver personalised movies.)
Funding	Industrial Project
Coordinator	Idiap Research Institute
Duration	2015.07.01 – 2016.02.29
Partner(s)	iRewind

Name	**GOOGLE-MOBILE** (Mobile Face and Voice Anti Spoofing)
Funding	Industrial Project
Coordinator	Idiap Research Institute
Duration	2015.05.15 – 2016.03.31

PROJECTS AWARDED IN 2016 AND STARTING IN 2017

Name	**SIMPLICID** (Seemless Customer Identification)
Funding	CTI
Coordinator	Idiap Research Institute
Duration	2017.03.01 – 2018.11.30
Partner(s)	Swisscard AECS GmbH

Name	**UNICITY** (3D scene understanding through machine learning to secure entrance zones)
Funding	CTI
Coordinator	Idiap Research Institute
Duration	2017.03.01 – 2018.11.30
Partner(s)	HES-SO Fribourg, FASTCOM Technology SA

Name	**COBALT** (Content Based Call Filtering)
Funding	Fondation The Ark
Coordinator	Idiap Research Institute
Duration	2017.03.01 – 2018.02.28
Partner(s)	Katia SA

Name	**RECOMEDIND** (RecoMed Industrialisation)
Funding	Fondation The Ark
Coordinator	recapp
Duration	2017.03.01 – 2017.07.31
Partner(s)	Idiap Research Institute, Clinique romande de réadaptation

Name	**FLOSS** (Flexible Linguistically-guided Objective Speech aSessment)
Funding	Hasler Foundation
Coordinator	Idiap Research Institute
Duration	2017.03.01 – 2020.02.2

Name	**CCSUMMARY** (Classification and Summarization of Written and Spoken Interactions with Customers)
Funding	IND – Industrial Project
Coordinator	Idiap Research Institute
Duration	2017.03.01 – 2020.02.29

Name	**ISUL** (Importance sampling for large-scale unsupervised learning)
Funding	SNSF Division II
Coordinator	Idiap Research Institute
Duration	2017.03.01 – 2019.02.28

Name	**MASS** (Multilingual Affective Speech Synthesis)
Funding	SNSF Division II
Coordinator	Idiap Research Institute
Duration	2017.05.01 – 2020.04.30

Books
and Book Chapters

[1] O. Aran, I. Ari, A. Kindiroglu, P. Santemiz, and L. Akarun, "Otomatik i̧saret dili tanıma ve türk i̧saret dili için bilgisayar uygulamaları," in Ellerle Konusmak: Türk ˙I̧saret Dili Ara̧stırmaları / Research on Turkish Sign Language, in Turkish, Koc University Press, 2016, pp. 471–498.

[2] A. G. Billard, S. Calinon, and R. Dillmann, "Learning from humans," in Handbook of Robotics, B. Siciliano and O. Khatib, Eds., 2nd Edition, Secaucus, NJ, USA: Springer, 2016, ch. 74, pp. 1995– 2014.

[3] I. Chingovska, N. Erdogmus, A. Anjos, and S. Marcel, "Face recognition systems under spoofing attacks," in Face Recognition Systems Under Spoofing Attacks, 1st, Springer International Publishing, Feb. 2016, ch. 8, pp. 165–194.

[4] D. Gatica-Perez, S. Ruiz-Correa, and D. Santani, "What tripadvisor can't tell: Crowdsourcing urban impressions for whole cities," in Digital Polis, A. d. Biase, N. Ottaviano, and O. Zaza, Eds., L'Oeil d'Or (translated to French.), 2016.

[5] D. Ginsbourger, J. Baccou, C. Chevalier, and F. Perales, "Design of computer experiments using competing distances between set-valued inputs," in MODa 11 - Advances in Model-Oriented Design and Analysis, ser. Contributions to Statistics, Springer International Publishing, 2016, pp. 123–131.

[6] D. Ginsbourger, O. Roustant, D. Schuhmacher, N. Durrande, and N. Lenz, "On anova decompositions of kernels and gaussian random field paths," in Monte Carlo and Quasi-Monte Carlo Methods, ser. Springer Proceedings in Mathematics & Statistics, vol. 163, Springer International Publishing, 2016, pp. 315–330.

[7] M. Günther, L. El Shafey, and S. Marcel, "Face recognition in challenging environments: An experimental and reproducible research survey," in Face Recognition Across the Imaging Spectrum, T. Bourlai, Ed., 1st ed., Springer, Feb. 2016.

Articles in
Scientific Journals

[1] A. Asaei, H. Bourlard, M. J. Taghizadeh, and V. Cevher, "Computational methods for underdetermined convolutive speech localization and separation via model-based sparse component analysis," Speech Communication, vol. 76, pp. 201–217, 2016.

[2] A. Asaei, M. J. Taghizadeh, S. Haghighatshoar, B. Raj, H. Bourlard, and V. Cevher, "Binary sparse coding of convolutive mixtures for sound localization and separation via spatialization," IEEE Transactions on Signal Processing, vol. 64, no. 3, pp. 567–579, 2016.

[3] U. Avci and O. Aran, "Predicting the performance in decision-making tasks: From individual cues to group interaction," IEEE Transactions on Multimedia, vol. 18, no. 4, pp. 643–658, 2016.

[4] D. Azzimonti, J. Bect, C. Chevalier, and D. Ginsbourger, "Quantifying uncertainties on excursion sets under a gaussian random field prior," SIAM/ASA J. Uncertainty Quantification, vol. 4, no. 1, pp. 850–874, 2016.

[5] S. Calinon, "A tutorial on task-parameterized movement learning and retrieval," Intelligent Service Robotics, vol. 9, no. 1, pp. 1–29, Jan. 2016.

[6] G. Can, J.-M. Odobez, and D. Gatica-Perez, "Evaluating shape representations for maya glyph classification," ACM Journal on Computing and Cultural Heritage (JOCCH), vol. 9, no. 3, Sep. 2016.

[7] A. Cerekovic, O. Aran, and D. Gatica-Perez, "Rapport with virtual agents: What do human social cues and personality explain?" IEEE Transactions on Affective Computing, 2016.

LIST OF PUBLICATIONS IN 2016

[8] M. Cernak, A. Asaei, and H. Bourlard, "On structured sparsity of phonological posteriors for linguistic parsing," Speech Communication, vol. 84, pp. 36–45, 2016.

[9] M. Cernak, Š. Beˇnuš, and A. Lazaridis, "Speech vocoding for laboratory phonology," Computer Speech and Language, L. ten Bosch, Ed., 2016.

[10] M. Cernak, A. Lazaridis, A. Asaei, and P. N. Garner, "Composition of deep and spiking neural networks for very low bit rate speech coding," IEEE/ACM Trans. on Audio, Speech and Language Processing, 2016.

[11] K. G. Chan, S. J. Streichan, L. A. Trinh, and M. Liebling, "Simultaneous temporal superresolution and denoising for cardiac fluorescence microscopy," IEEE Transactions on Computational Imaging, 2016, in press.

[12] P. Dighe, A. Asaei, and H. Bourlard, "Sparse modeling of neural network posterior probabilities for exemplar-based speech recognition," Speech Communication: Special Issue on Advances in Sparse Modeling and Low-rank Modeling for Speech Processing, vol. 76, 230–244, Feb. 2016.

[13] M. Ferras, S. Madikeri, and H. Bourlard, "Speaker diarization and linking of meeting data," IEEE/ ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 11, pp. 1935–1945, Nov. 2016.

[14] M. Ferras, S. Madikeri, S. Dey, P. Motlicek, and H. Bourlard, "A large-scale open-source acoustic simulator for speaker recognition," IEEE Signal Processing Letters, vol. 23, no. 4, pp. 527 –531, 2016.

[15] M. Focchi, A. del Prete, I. Havoutis, R. Featherstone, D. G. Caldwell, and C. Semini, "High-slope terrain locomotion for torque-controlled quadruped robots," Autonomous Robots, 2016.

[16] K. A. Funes Mora and J.-M. Odobez, "Gaze estimation in the 3d space using rgb-d sensors. towards head-pose and user invariance.," International Journal of Computer Vision, vol. 118, no. 2, pp. 194–216, Jun. 2016, First online: 13 November 2015.

[17] L. Fusco, R. Lefort, K. C. Smith, F. Benmansour, G. Gonzalez, C. Barilari, B. Rinn, F. Fleuret, P. Fua, and O. Pertz, "Computer vision profiling of neurite outgrowth dynamics reveals spatiotemporal modularity of rho gtpase signaling," Journal of Cell Biology, vol. 212, no. 1, pp. 91–111, Jan. 2016.

[18] P. Gay, S. Meignier, P. Deleglise, and J.-M. Odobez, "Crf-based context modeling for person identification in broadcast videos," Frontiers in ICT: Computer Image Analysis, vol. 3, 2016.

[19] D. Ginsbourger, O. Roustant, and N. Durrande, "On degeneracy and invariances of random fields paths with applications in gaussian process modelling," Journal of Statistical Planning and Inference, vol. 170, pp. 117–128, Mar. 2016.

[20] M. Habibi, P. Mahdabi, and A. Popescu-Belis, "Question answering in conversations: Query refinement using contextual and semantic information," Data & Knowledge Engineering Journal, Jun. 2016.

[21] I. Himawan, P. Motlicek, D. Imseng, and S. Sridharan, "Feature mapping using far-field microphones for distant speech recognition," Speech Communication, vol. 83, pp. 1–9, Oct. 2016, A publication of the European Association for Signal Processing (EURASIP) and of the International Speech Communication Association (ISCA).

[22] R. Hu, J.-M. Odobez, and D. Gatica-Perez, "Extracting maya glyphs from degraded ancient documents via image segmentation," Journal on Computing and Cultural Heritage, 2016.

[23] I. Kuzborskij and F. Orabona, "Fast rates by transferring from auxiliary hypotheses," Machine Learning, 2016.

[24] I. Kuzborskij, F. Orabona, and B. Caputo, "Scalable greedy algorithms for transfer learning," Computer Vision and Image Understanding, 2016.

[25] L. Lefakis and F. Fleuret, "Jointly informative feature selection," Journal of Machine Learning Research, 2016.

[26] A. Morales, J. Fierrez, R. Tolosana, J. Ortega-Garcia, J. Galbally, M. Gomez-Barrero, A. Anjos, and S. Marcel, "Keystroke biometrics ongoing competition," IEEE Access, vol. 4, pp. 7736–7746, Nov. 2016.

[27] L. S. Nguyen and D. Gatica-Perez, "Hirability in the wild: Analysis of online conversational video resumes," IEEE Trans. on Multimedia, vol. 18, no. 7, pp. 1422–1437, Jul. 2016.

[28] N. Pappas and A. Popescu-Belis, "Adaptive sentiment-aware one-class collaborative filtering," Expert Systems with Applications, vol. 43, pp. 23–41, Jan. 2016.

[29] R. Rasipuram and M. Magimai.-Doss, "Articulatory feature based continuous speech recognition using probabilistic lexical modeling," Computer Speech and Language, vol. 36, pp. 233–259, 2016.

[30] M. Razavi, R. Rasipuram, and M. Magimai.-Doss, "Acoustic data-driven grapheme-to-phoneme conversion in the probabilistic lexical modeling framework," Speech Communication, vol. 80, 2016.

[31] L. Rozo, S. Calinon, D. G. Caldwell, P. Jimenez, and C. Torras, "Learning physical collaborative robot behaviors from human demonstrations," IEEE Trans. on Robotics, vol. 32, no. 3, pp. 513– 527, Jun. 2016.

[32] L. Rozo, J. Silverio, S. Calinon, and D. G. Caldwell, "Learning controllers for reactive and proactive behaviors in human-robot collaboration," Frontiers in Robotics and AI, vol. 3, no. 30, pp. 1–11, Jun. 2016.

[33] S. Ruiz-Correa, D. Santani, B. Ramirez Salazar, I. Ruiz-Correa, F. Rendon-Huerta, C. Olmos Carrillo, B. Sandoval Mexicano, A. Arcos Garcia, R. Hasimoto-Beltran, and D. Gatica-Perez, "Sensecityvity: Mobile crowdsourcing, urban awareness, and collective action in mexico," IEEE Pervasive Computing, 2016.

[34] S. Ruiz-Correa, D. Santani, B. Ramirez Salazar, I. Ruiz-Correa, F. Rendon-Huerta, C. Olmos Carrillo, B. Sandoval Mexicano, A. Arcos Garcia, R. Hasimoto-Beltran, and D. Gatica-Perez, "Sensecityvity: Mobile sensing, urban awareness, and collective action in mexico," IEEE Pervasive Computing, 2016.

[35] N. Suditu and F. Fleuret, "Adaptive relevance feedback for large-scale image retrieval," Multimedia Tools and Applications, vol. 75, no. 12, pp. 6777–6807, Jun. 2016.

[36] A. K. Tanwani and S. Calinon, "Learning robot manipulation tasks with task-parameterized semitied hidden semi-markov model," IEEE Robotics and Automation Letters, vol. 1, no. 1, D. Kulic and D. Lee, Eds., pp. 235–242, Jan. 2016.

[37] R. Ullmann and H. Bourlard, "Predicting the intrusiveness of noise through sparse coding with auditory kernels," Speech Communication, vol. 76, pp. 186–200, 2016.

[38] J. Velasco, D. Pizarro, J. Macias-Guarasa, and A. Asaei, "Tdoa matrices: Algebraic properties and their application to robust denoising with missing data," IEEE Transactions on Signal Processing, vol. 64, no. 20, pp. 5242–5254, Jun. 2016.

[39] X.Wang, E. Turetken, F. Fleuret, and P. Fua, "Tracking interacting objects using intertwined flows," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016.

[40] D. Wu, L. Pigou, P.-J. Kindermans, N. Le, L. Shao, J. Dambre, and J.-M. Odobez, "Deep dynamic neural networks for multimodal gesture segmentation and recognition,"IEEE Transactions on Pattern Analysis and Machine Intelligence

PhD Theses

[1] R. Lebret, " Building word embeddings for solving natural language processing, " Thèse EPFL, nr 7148, PhD thesis, École Polytechnique Fédérale de Lausanne, Sep. 2016.

[2] J. Legrand, " Word sequence modeling using deep learning: And end-to-end approach and its applications, " PhD thesis, EPFL, Oct. 2016.

[3] D. Palaz, " Towards end-to-end speech recognition, " Thèse EPFL nr 7054, PhD thesis, Ecole polytechnique Fédérale de Lausanne, Jul. 2016.

[4] N. Pappas, " Learning explainable user sentiment and preferences for information filtering, " Thèse EPFL, nr 6920, PhD thesis, École Polytechnique Fédérale de Lausanne, Mar. 2016.

[5] D. Santani, " Computational analysis of urban places using mobile crowdsensing, " These EPFL, nr 7243, PhD thesis, Ecole Polytechnique Federale de Lausanne, Nov. 2016.

[6] R. Ullmann, " Can you hear me now? automatic assessment of background noise intrusiveness and speech intelligibility in telecommunications, " PhD thesis, Sciences et Techniques de l'Ingénieur (STI), Jun. 2016.

Articles in Conference Proceedings

[1] O. Aran, D. Sanchez-Cortes, T.-M.-T. Do, and D. Gatica-Perez, " Anomaly detection in elderly daily behavior in ambient sensing environments, " in Proceedings of the 7th Int. Workshop on Human Behavior Understanding, ACM Multimedia, 2016, Amsterdam, Netherlands, 2016.

[2] A. Asaei, M. Cernak, and M. Laganaro, " Paos markers: Trajectory analysis of selective phonological posteriors for assessment of progressive apraxia of speech, " in Proceeding on the 7th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT), 2016.

[3] A. Asaei, G. Luyet, M. Cernak, and H. Bourlard, " Efficient posterior exemplar search space hashing exploiting class-specific sparsity structures, " in Interspeech, San Francisco, CA, 2016.

[4] P. Baqué, T. Bagautdinov, F. Fleuret, and P. Fua, " Principled parallel mean-field inference for discrete random fields, " in Proceedings of the IEEE international conference on Computer Vision and Pattern Recognition, 2016.

[5] K. Bascol, R. Emonet, E. Fromont, and J.-M. Odobez, " Unsupervised interpretable pattern discovery in time series using autoencoders, " in IAPR Int. Workshops on Structural and Syntactic Pattern Recognition (SSPR), Nov. 2016.

[6] D. Berio, S. Calinon, and F. F. Leymarie, " Learning dynamic graffiti strokes with a compliant robot, " in Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems, Oct. 2016, pp. 3981–3986.

[7] C. A. Bhatt, A. Popescu-Belis, and M. Cooper, " Audiovisual summarization of lectures and meetings using a segment similarity graph, " in Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR), ACM, New York, NY: ACM Press, 2016.

[8] S. Calinon, "Stochastic learning and control in multiple coordinate systems," in Intl Workshop on Human-Friendly Robotics, Genoa, Italy, 2016, pp. 1–5.

[9] G. Can, J.-M. Odobez, C. Pallan Gayol, and D. Gatica-Perez, "Ancient maya writings as highdimensional data: A visualization approach," in Digital Humanities (DH), Krakow, Jul. 2016.

[10] O. Canévet and F. Fleuret, "Large scale hard sample mining with monte carlo tree search," in Proceedings of the IEEE international conference on Computer Vision and Pattern Recognition, 2016.

[11] O. Canévet, C. Jose, and F. Fleuret, "Importance sampling tree for large-scale empirical expectation," in Proceedings of the International Conference on Machine Learning (ICML), New-York, Jun. 2016.

[12] M. Cernak, A. Asaei, P.-E. Honnet, P. N. Garner, and H. Bourlard, "Sound pattern matching for automatic pro-sodic event detection," in Interspeech, San Francisco, USA, Sep. 2016.

[13] M. Cernak and P. N. Garner, "Phonvoc: A phonetic and phonological vocoding toolkit," in Interspeech, San Francisco, USA, Sep. 2016.

[14] K. G. Chan and M. Liebling, "A point-spread-func-tion-aware filtered backprojection algorithm for focal-plane-scanning optical projection tomography," in 2016 IEEE International Symposium on Biomedical Imaging, Apr. 2016.

[15] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, and S. Marcel, "The replay-mobile face presentation-at-tack database," in Proceedings of the International Conference on Biometrics Special Interests Group, Sep. 2016.

[16] N. Dawalatabad, S. Madikeri, C. C. Sekhar, and H. A. Murthy, "Two-pass ib based speaker diarization system using meeting-specific ann based features," in Proceedings of Interspeech 2016, Sep. 2016, pp. 2199–2203.

[17] S. Dey, S. Madikeri, M. Ferras, and P. Motlicek, "Deep neural network based posteriors for text-dependent speaker verification," in Proceedings of 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2016), Shanghai: IEEE, Mar. 2016, pp. 5050– 5054.

[18] S. Dey, S. Madikeri, and P. Motlicek, "Information theo-retic clustering for unsupervised domainadaptation," in Proceedings of 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2016), Shanghai: IEEE, Mar. 2016, pp. 5580–5584.

[19] P. Dighe, G. Luyet, A. Asaei, and H. Bourlard, "Exploiting low-dimensional structures to enhance dnn based acoustic modeling in speech recognition," in Proceedings of 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2016), Shanghai: IEEE, Mar. 2016, pp. 5690–5694.

[20] M. Ferras, S. Madikeri, S. Dey, P. Motlicek, and H. Bourlard, "Inter-task system fusion for speaker recogni-tion," in Proceeedings of the INTERSPEECH, 2016.

[21] M. Ferras, S. Madikeri, P. Motlicek, and H. Bourlard, "System fusion and speaker linking for longitudinal diarization of tv shows," in Proceedings of 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2016), Shanghai: IEEE, Mar. 2016, pp. 5495– 5499.

[22] A. Finnerty, S. Muralidhar, L. S. Nguyen, F. Pianesi, and D. Gatica-Perez, "Stressful first impressions in job inter-views," in Proceedings of the 18th ACM International Conference on Multimodal Interaction, Nov. 2016, pp. 325–332.

[23] T. de Freitas Pereira and S. Marcel, "Heterogeneous face recognition using inter-session variability mod-elling," in IEEE Computer Society Workshop on Biometrics, Las Vegas – USA: IEEE, Jun. 2016.

[24] T. Gabor Csapo, G. Nemeth, M. Cernak, and P. N. Garner, "Modeling unvoiced sounds in statistical para-metric speech synthesis with a continuous vocoder," in Proc. of EUSIPCO, Budapest, Hungary, 2016.

LIST OF PUBLICATIONS IN 2016

[25] J. Gancet, P. Weiss, G. Antonelli, M. F. Pfingsthorn, S. Calinon, A. Turetta, C. Walen, D. Urbina, S. Govindaraj, P. Letier, X. Martinez, J. Salini, B. Chemisky, G. Indiveri, G. Casalino, P. Di Lillo, E. Simetti, D. De Palma, A. Birk, A. K. Tanwani, I. Havoutis, A. Caffaz, and L. Guilpain, "Dexterous undersea interventions with far distance onshore supervision: The dexrov project," in IFAC Conference on Control Applications in Marine Systems (CAMS), Trondheim, Norway, Sep. 2016, pp. 414–419.

[26] B. Gerazov and P. N. Garner, "An agonist-antagonist pitch production model," in Lecture Notes in Artificial Intelligence: 18th International Conference, SPECOM 2016, vol. 9811, Budapest, Hungary, Aug. 2016, pp. 84–91.

[27] B. Gerazov, A. Gjoreski, A. Melov, P.-E. Honnet, Z. Ivanovski, and P. N. Garner, "Unified prosody model based on atom decomposition for emphasis detection," in Proceedings of ETAI, Struga, Macedonia, Sep. 2016.

[28] J.-P. Goldman, P.-E. Honnet, R. Clark, P. N. Garner, M. Ivanova, A. Lazaridis, H. Liang, T. Macedo, B. Pfister, M. S. Ribeiro, E. Wehrli, and J. Yamagishi, "The siwis database: A multilingual speech database with acted emphasis," in Proceedings of Interspeech, San Francisco, USA, Sep. 2016, pp. 1532–1535.

[29] L. Guillou, C. Hardmeier, P. Nakov, S. Stymne, J. Tiedemann, Y. Versley, M. Cettolo, B. Webber, and A. Popescu-Belis, "Findings of the 2016 wmt shared task on cross-lingual pronoun prediction," in Proceedings of WMT 2016 (First Conference on Machine Translation), Association for Computational Linguistics, Berlin, Germany, 2016, 525–542.

[30] I. Havoutis and S. Calinon, "Learning assistive tele-operation behaviors from demonstration," in Proc. IEEE International Symposium on Safety, Security and Rescue Robotics, Oct. 2016, pp. 258–263.

[31] P.-E. Honnet and P. N. Garner, "Emphasis recreation for tts using intonation atoms," in 9th ISCA Speech Synthesis Workshop, Sunnyvale, CA, USA, Sep. 2016, pp. 14–20.

[32] R. Hu, J.-M. Odobez, and D. Gatica-Perez, "Assessing a shape descriptor for analysis of mesoamerican hieroglyphics: A view towards practice in digital humanities," in Digital Humanities Conference (DH), Krakow, Jul. 2016.

[33] C. Jose and F. Fleuret, "Scalable metric learning via weighted approximate rank component analysis," in ECCV 2016, 2016.

[34] P. Korshunov and S. Marcel, "Cross-database evaluation of audio-based spoofing detection systems," in Interspeech, San Francisco, USA, Sep. 2016.

[35] P. Korshunov and S. Marcel, "Joint operation of voice biometrics and presentation attack detection," in IEEE International Conference on Biometrics: Theory, Applications and Systems, Niagara Falls, Buffalo New York, USA, Sep. 2016.

[36] P. Korshunov, S. Marcel, H. Muckenhirn, A. R. Gonçalves, A. G. S. Mello, R. P. V. Violato, F. O. Simões, M. U. Neto, M. de Assis Angeloni, J. A. Stuchi, H. Dinkel, N. Chen, Y. Qian, D. Paul, G. Saha, and M. Sahidullah, "Overview of btas 2016 speaker anti-spoofing competition," in IEEE International Conference on Biometrics: Theory, Applications and Systems, Niagara Falls, NY, USA, Sep. 2016.

[37] I. Kuzborskij, F. M. Carlucci, and B. Caputo, "When naïve bayes nearest neighbors meet convolutional neural networks," in Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, Jun. 2016.

[38] A. Lazaridis, M. Cernak, and P. N. Garner, "Probabilistic amplitude demodulation features in speech synthesis for improving prosody," in Proceedings of Interspeech, San Francisco, USA, 2016.

[39] A. Lazaridis, M. Cernak, P.-E. Honnet, and P. N. Garner, "Investigating spectral amplitude modulation phase hierarchy features in speech synthesis," in 9th ISCA Speech Synthesis Workshop, Sunnyvale, CA, USA, Sep. 2016.

LIST OF PUBLICATIONS IN 2016

[40] A. Lazaridis, I. Himawan, P. Motlicek, I. Mporas, and P. N. Garner, "Investigating cross-lingual multi-level adaptive networks: The importance of the correlation of source and target languages," in Proceedings of the International Workshop on Spoken Language Translation, Seattle, WA, USA, Dec. 2016.

[41] N. Le, A. Heili, and J.-M. Odobez, "Long-term time-sensitive costs for crf-based tracking by detection," in 2ndWorkshop on Benchmarking Multi-target Tracking: MOTChallenge 2016, Amsterdam, Oct. 2016.

[42] N. Le, A. Heili, D. Wu, and J.-M. Odobez, "Temporally subsampled detection for accurate and efficient face tracking and diarization," in International Conference on Pattern Recognition, Cancun, Mexico: IEEE, Dec. 2016.

[43] N. Le, S. Meignier, and J.-M. Odobez, "Eumssi team at the mediaeval person discovery challenge 2016," in MediaEval Benchmarking Initiative for Multimedia Evaluation, Hilversum, Netherlands, Oct. 2016.

[44] N. Le and J.-M. Odobez, "Learning multimodal temporal representation for dubbing detection in broadcast media," in ACM Multimedia, Amsterdam: ACM, Oct. 2016.

[45] R. Lebret, "Building word embeddings for solving natural language processing," Thèse EPFL, nr 7148, PhD thesis, École Polytechnique Fédérale de Lausanne, Jul. 2016.

[46] J. Legrand, "Word sequence modeling using deep learning: And end-to-end approach and its applications," PhD thesis, EPFL, Aug. 2016.

[47] J. Legrand, M. Auli, and R. Collobert, "Neural network-based word alignment through score aggregation," in Proceedings of the ACL 1st Conference on Machine Translation, 2016.

[48] J. Legrand and R. Collobert, "Deep neural networks for syntactic parsing of morphologically rich languages," in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016.

[49] J. Legrand and R. Collobert, "Phrase representations for multiword expressions," in Proceedings of the 12th Workshop on Multiword Expressions, 2016.

[50] J. Liyanapathirana and A. Popescu-Belis, "Using the ted talks to evaluate spoken post-editing of machine translation," in Proceedings of the 10th Language Resources and Evaluation Conference (LREC), Portoroz, Slovenia, 2016.

[51] N.-Q. Luong and A. Popescu-Belis, "A contextual language model to improve machine translation of pronouns by re-ranking translation hypotheses," in European Association for Machine Translation, May 2016.

[52] N.-Q. Luong and A. Popescu-Belis, "Improving pronoun translation by modeling coreference uncertainty," in Proceedings of the First Conference on Machine Translation (WMT16), Berlin, Germany, Aug. 2016.

[53] N.-Q. Luong and A. Popescu-Belis, "Pronoun language model and grammatical heuristics for aiding pronoun prediction," in Proceedings of the First Conference on Machine Translation (WMT16), Berlin, Germany: ACL, Aug. 2016.

[54] G. Luyet, P. Dighe, A. Asaei, and H. Bourlard, "Low-rank representation of nearest neighbor phone posterior probabilities to enhance dnn acoustic modeling," in Interspeech, 2016.

[55] P. Mahdabi and A. Popescu-Belis, "Comparing two strategies for query expansion in a news monitoring system," in Natural Language Processing and Information Systems: 21st International Conference on Applications of Natural Language to Information Systems, NLDB 2016, ser. Lecture Notes in Computer Science, vol. 9612, Springer-Verlag, 2016, pp. 267–275.

[56] C. Mastalli, I. Havoutis, M. Focchi, C. Semini, and D. G. Caldwell, "Hierarchical planning of dynamic movements without scheduled contact sequences," in Proceedings of the IEEE International Conference of Robotics and Automation, 2016.

[57] H. Muckenhirn, M. Magimai.-Doss, and S. Marcel, "Presentation attack detection using long-term spectral statistics for trustworthy speaker verification," in Proceedings of International Conference of the Biometrics Special Interest Group (BIOSIG), Sep. 2016.

[58] S. Muralidhar, J. M. R. Costa, L. S. Nguyen, and D. Gatica-Perez, "Dites-moi: Wearable feedback on conversational behavior," in Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia, Dec. 2016.

[59] S. Muralidhar, L. S. Nguyen, D. Frauendorfer, J.-M. Odobez, M. Schmid Mast, and D. Gatica- Perez, "Training on the job: Behavioral analysis of job interviews in hospitality," in Proceedings of the 18th ACM International Conference on Multimodal Interaction, Nov. 2016, pp. 84–91.

[60] J. Newling and F. Fleuret, "Fast k-means with accurate bounds," in Proceedings of the International Conference on Machine Learning (ICML), New York, 2016.

[61] J. Newling and F. Fleuret, "Nested mini-batch k-means," in Proceedings of NIPS, 2016.

[62] C. Oertel, J. D. Lopes, Y. Yu, K. A. Funes Mora, J. Gustafson, A. Black, and J.-M. Odobez, "Towards building an attentive artificial listener: On the perception of attentiveness in audio-visual feedback tokens," in Proceedings of the 18th ACM International Conference on Multimodal Interaction, Tokyo, Japan: ACM, Nov. 2016, pp. 21–28.

[63] D. Palaz, "Towards end-to-end speech recognition," Thèse EPFL nr 7054, PhD thesis, Ecole polytechnique Fédérale de Lausanne, 2016.

[64] N. Pappas, "Learning explainable user sentiment and preferences for information filtering," Thèse EPFL, nr 6920, PhD thesis, École Polytechnique Fédérale de Lausanne, Mar. 2016.

[65] N. Pappas and A. Popescu-Belis, "Human versus machine attention in document classification: A dataset with crowdsourced annotations," in Proceedings of the EMNLP 2016 Workshop on Natural Language Processing for Social Media, Austin, USA, 2016.

[66] N. Pappas, M. Topkara, M. Redi, B. Jou, T. Chen, H. Liu, and S.-F. Chang, "Multilingual visual sentiment concept matching," in International Conference on Multimedia Retrieval, 2016.

[67] P. H. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollar, "Learning to refine object segments," in Computer Vision – ECCV 2016, ser. Lecture Notes in Computer Science, vol. 9905, Amsterdam: Springer, Sep. 2016, pp. 75–91.

[68] A. Popescu-Belis, "Manual and automatic labeling of discourse connectives for machine translation (keynote paper)," in TextLink: Structuring Discourse in Multilingual Europe (Handbook of the Second Action Conference), Budapest, Hungary, 2016, pp. 16–20.

[69] X. Pu, M. A. Chatti, H. Thues, and U. Schroeder, "Wiki-lda: A mixed-method approach for effective interest mining on twitter data," in Proceedings of CSEDU 2016, Apr. 2016.

[70] D. Ram, A. Asaei, and H. Bourlard, "Subspace detection of dnn posterior probabilities via sparse representation for query by example spoken term detection," in Interspeech, 2016.

[71] R. Rasipuram, M. Cernak, and M. Magimai.-Doss, "Hmm-based non-native accent assessment using posterior features," in Proceedings of Interspeech, San Francisco, USA, Sep. 2016.

[72] M. Razavi and M. Magimai.-Doss, "Improving under-resourced language asr through latent subword unit space discovery," in Proceedings of Interspeech, 2016.

LIST OF PUBLICATIONS IN 2016

[73] E. Roman-Rangel, G. Can, S. Marchand-Maillet, R. Hu, C. Pallan Gayol, G. Krempel, J. Spotak, J.-M. Odobez, and D. Gatica-Perez, " Transferring neural representations for low-dimensional indexing of maya hieroglyphic art, " in Proc. ECCV Workshop on Computer Vision for Art Analysis, Amsterdam, Oct. 2016.

[74] D. Santani, " Computational analysis of urban places using mobile crowdsensing, " These EPFL, nr 7243, PhD thesis, Ecole Polytechnique Federale de Lausanne, Sep. 2016.

[75] D. Santani, J.-I. Biel, F. Labhart, J. Truong, S. Landolt, E. Kuntsche, and D. Gatica-Perez, " The night is young: Urban crowdsourcing of nightlife patterns, " in Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ser. UbiComp '16, ACM, Sep. 2016.

[76] D. Santani, R. Hu, and D. Gatica-Perez, " Innerview: Learning place ambiance from social media images, " in Proceedings of the 24th ACM International Conference on Multimedia, ser. MM '16, ACM, Oct. 2016.

[77] M. Secujski, B. Gerazov, T. Gabor Csapo, V. Delic, P. N. Garner, A. Gjoreski, D. Guennec, Z. Ivanovski, A. Melov, G. Nemeth, A. Stojkovi'c, and G. Szaszak, " Design of a speech corpus for research on cross-lingual prosody transfer, " in Lecture Notes in Artificial Intelligence: 18th International Conference, SPECOM 2016, vol. 9811, Budapest, Hungary, Aug. 2016, pp. 199–206.

[78] A. K. Tanwani and S. Calinon, " Online inference in bayesian non-parametric mixture models under small variance asymptotics, " in NIPS workshop on Advances in Approximate Bayesian Inference, Barcelona, Spain, 2016, pp. 1–5.

[79] R. Ullmann, " "can you hear me now?" — automatic assessment of background noise intrusiveness and speech intelligibility in telecommunications, " PhD thesis, Sciences et Techniques de l'Ingénieur (STI), 2016.

[80] S. Zagoruyko, A. Lerer, T.-Y. Lin, P. H. O. Pinheiro, S. Gross, S. Chintala, and P. Dollar, " A multipath network for object detection, " in Proceedings of the British Machine Vision Conference, BMVA Press, Sep. 2016.

[81] M. Zeestraten, S. Calinon, and D. G. Caldwell, " Variable duration movement encoding with minimal intervention control, " in Proc. of the IEEE Intl Conf. on Robotics and Automation (ICRA), May 2016, pp. 497–503.

[82] M. Zeestraten, A. Pereira, M. Althoff, and S. Calinon, " Online motion synthesis with minimal intervention control and formal safety guarantees, " in Proc. IEEE Intl Conf. on Systems, Man, and Cybernetics, Budapest, Hungary, Oct. 2016.